Higher Order Fuzzy Rule Interpolation

Chengyuan Chen

Supervisors: Prof. Qiang Shen
Dr. Richard Jensen

Ph.D. Thesis

Department of Computer Science
Institute of Mathematics, Physics and Computer Science
Aberystwyth University

May 20, 2015

Declaration and Statement

DECLARATION

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed .... (candidate)

Date ....

STATEMENT 1

This thesis is the result of my own investigations, except where otherwise stated. Where **correction services**[1] have been used, the extent and nature of the correction is clearly marked in a footnote(s).

Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed .... (candidate)

Date ....

STATEMENT 2

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed .... (candidate)

Date ....

[1] This refers to the extent to which the text has been corrected by others.

Abstract

Fuzzy inference is an effective means for representing and handling vagueness and imprecision. As a particular type of fuzzy inference, fuzzy rule interpolation enhances the performance of the inference when a given observation has no overlap with the antecedent values of any of the existing rules. In such cases, conventional fuzzy inference methods cannot derive a conclusion, but fuzzy rule interpolation methods can still obtain a certain conclusion. Unfortunately, very little of the existing work on fuzzy rule interpolation can conjunctively handle more than one form of uncertainty in the rules or observations. In particular, the difficulty in defining the required precise-valued membership functions for the fuzzy sets that are used by conventional fuzzy rule interpolation techniques significantly restricts their application.

In this thesis, a novel framework termed "higher order fuzzy rule interpolation" is proposed in an attempt to address such difficulties. The proposed framework allows the representation, handling and utilisation of different types of uncertainty in knowledge. This allows transformation-based fuzzy rule interpolation techniques to harness and utilise the additional uncertainty in order to implement a fuzzy interpolative reasoning system. Final conclusions can then be derived by performing higher order interpolation over this representation.

The techniques for the representation and handling of uncertainty are organised in this framework such that in circumstances when different types of uncertainty are encountered the inference process can deal with them in an appropriate way. A rough-fuzzy set based rule interpolation approach is proposed in this work, by exploiting the concept of rough-fuzzy sets and generalising scale and move transformation-based fuzzy interpolation. A type-2 fuzzy set based interpolation approach is also presented as an alternative implementation of the framework. The effectiveness of this work in improving the robustness of fuzzy rule interpolation is demonstrated through the practical application to the prediction of disease rates in remote villages. Moreover, this framework is also further evaluated with application to other realistic decision making problems. The resultant accuracy reveals the efficacy of this research.

Acknowledgements

Foremost, I would like to express my deepest gratitude and thanks to my supervisors: Prof. Qiang Shen and Dr. Richard Jensen, for their patience, encouragement, and guidance throughout this research. I have enjoyed the opportunity to learn from their knowledge and experience.

I must also sincerely thank Dr. Neil S. Mac Parthaláin. His kind assistance and helpful advice are always appreciated.

My sincere gratitude goes to my parents, my wife, and her parents. I would not have completed this work without their love, support, and understanding over the last few years.

I am grateful to all of my friends in the Advanced Reasoning Group. I have enjoyed many useful and entertaining discussions with them.

I would like to thank the staff in the Department of Computer Science who have helped me during my study.

Thanks also goes to the colleagues from Chongqing University of Science and Technology for their support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In conventional (or hard) computing such as boolean logic, binary systems, numerical analysis and crisp software, the prime considerations are rigour, precision and certainty. However, inexact information involving imprecision and uncertainty exists ubiquitously in the real world, so that such information is difficult to process using hard computing. In contrast, as an emerging collection of methodologies and techniques, soft computing (SC) [12, 202, 219] exploits the tolerance for imprecision, uncertainty, partial truth and approximations in order to achieve close resemblance with human activity and reasoning intuition. Unlike hard computing, SC aims to represent the ambiguity in human thinking with real life uncertainty [116, 117]. Therefore, SC can be particularly beneficial for problems which are too complex to be directly handled by human beings.

The core methodologies of SC are fuzzy logic [104, 217], neural networks [10, 68], evolutionary computation [4, 46], and probabilistic reasoning [145, 155]. Each of these foundations provides solutions with complementary reasoning and search methods for real-world problems. In particular, fuzzy logic is primarily concerned with imprecision. Its main contribution is a methodology for computing with words [113, 131, 134, 220], providing foundations for approximate reasoning [63, 216] using imprecise propositions based on fuzzy set (FS) theory [54, 212, 214, 226]. The importance of fuzzy logic derives from the fact that most modes of human reasoning, especially common sense reasoning, are approximate in nature [218, 221]. Fuzzy logic is useful for dealing with non-linear, uncertain and complex systems such as information processing and mechanical control [17, 21, 48, 112, 118, 130, 178].

This is usually implemented by fuzzy inference systems, which is the main topic of this thesis.

1.1 Fuzzy Inference Systems

A fuzzy inference system (FIS) is a way of formulating the mapping from given inputs to an output using fuzzy logic. The mapping then provides a basis from which decisions can be made, or patterns can be discerned. The concept of FISs is based on fuzzy logic, fuzzy IF-THEN rules [52] and fuzzy reasoning, which jointly enable modelling complex systems in a way naturally used by humans [114]. The general architecture of an FIS is well-known in the literature [49, 104, 180], consisting of four conceptual components: *fuzzifier*, *rule base*, *inference engine* and *defuzzifier*. The characteristic of each component will be explained later in Section 2.2.

With crisp inputs and output, an FIS implements a non-linear mapping from its inputs space to output space. This mapping is achieved by a number of fuzzy IF-THEN rules, each of which describes the local behaviour of the mapping. In particular, the antecedent of a rule represents a fuzzy region in the input space, while the consequence indicates the inferred consequent in the output region.

There are two ways to construct a fuzzy rule base for a given problem. The first class of FISs directly translates expert knowledge to fuzzy rules, so that these FISs are called fuzzy expert systems or fuzzy controllers [18, 128, 130]. Since rules are fuzzy representations of expert knowledge, these FISs offer a high semantic level and a good generalisation capability. However, the complexity of large systems may lead to an insufficient accuracy in the simulation results. Such drawback leads to the other class of FISs, which is a data-driven fuzzy system. The fuzzy rules are obtained from data by machine learning techniques rather than expert knowledge [139, 146, 179, 211].

1.2 Fuzzy Rule Interpolation

Given a fuzzy rule base generated in either of the above two ways, there are a number of fuzzy inference mechanisms, such as compositional rule of inference [214] and similarity-based fuzzy reasoning [31, 32, 173, 174, 210], that can be utilised for deriving a conclusion from a given observation. However, *dense* rule bases are compulsory for these methods. Briefly, a dense rule base is a rule base where the input universe of discourse is covered completely. Given such a rule base and an observation that is at least partially covered by the rule base, the conclusion can be inferred from certain rules that intersect with the observation. However, for the case where a fuzzy rule base (termed: *sparse* rule base [183]) contains "gaps", if a given observation has no overlap with the antecedent values of any rule, conventional fuzzy inference methods cannot derive a conclusion.

Fortunately, using fuzzy rule interpolation (FRI) [106, 107], certain useful conclusions may still be obtained. Moreover, with the help of FRI, the complexity of a rule base can be reduced by omitting fuzzy rules which may be approximated from their neighbouring rules. Despite these advantages, the application of traditional FRI methods may lead to abnormal fuzzy conclusions. One particular issue is that the convexity of the derived fuzzy values is not guaranteed [165, 205], but convexity is often a crucial requirement for fuzzy inference in order to attain better interpretability in the results.

In order to overcome such drawbacks, a number of significant extensions to the original FRI methods have been proposed in the literature, including [6, 23, 34, 35, 39, 73, 76, 77, 108, 169, 183, 203, 204, 206, 208]. In particular, the scale and move transformation-based FRI approach [75, 76, 77] (abbreviated to T-FRI hereafter) and its generalisation [163] can handle interpolation and extrapolation involving multiple fuzzy rules, with each rule consisting of multiple antecedents. Such work also guarantees the uniqueness, as well as the normality and convexity of the interpolated conclusion. This approach has recently been further enhanced with an adaptive mechanism such that performing appropriate chaining of fuzzy interpolative inferences is supported [206].

1.3 Uncertainty in Fuzziness

Conventional FS theory and the aforementioned FRI techniques provide a basic means for uncertainty interpretation and uncertainty treatment. However, there is little work in FRI that can handle uncertainty in fuzziness itself. This is because these approaches are implemented based on conventional FS representations [212]. Whilst membership functions (MFs) play an important role in defining FSs, it is sometimes extremely difficult, if not impossible, to precisely define such MFs. There may be different types of uncertainty in fuzzy rule-based systems that need to be captured [60]: (1) The linguistic variables that are used in the antecedents and consequences of the given rules may be indiscernible. (2) The meanings of the words representing the values of the underlying variables may be vague because words can mean different things to different people. (3) An object can belong to an FS with a degree, but that degree may itself be uncertain. (4) The generated rules may be inconsistent when personal views are provided from a group of experts. (5) Observations attainable by inexact knowledge may be noisy and therefore randomly distributed.

Most of these types of uncertainty can be difficult to deal with in order to determine the crisp MFs of the FSs used. For instance, certain weather conditions are considered *cold* by all people, but others may be considered as *cold* by only certain individuals. The MFs for different people may therefore be different, depending on their perception, preference, experience, etc. This is shown in Figure 1.1, where x and μ denote an element in a given concept and its corresponding membership value, respectively. That is, both similarities and differences may exist in defining a given perception. Therefore, the representation of a concept should satisfy the requirements of not only the imprecise description but also both the *common* perception and *individual* perception. In this case, the membership values of a conventional (aka., type-1) FS may not be adequately represented precisely.

1.4 Framework for Higher Order Uncertainty

In this case, different types of uncertainty may influence the determination of the crisp MFs and thereby have different effects upon the efficacy of FRI. When facing such a higher order uncertainty, which is the uncertainty of evaluation about uncertainty, a simple approach may be just ignoring this higher-level information. However, an obvious drawback of this is that substantial information may be lost from discarding such uncertain knowledge. This, in turn, may lead to unacceptable inference conclusions. Alternative representations are needed in order to better understand and manipulate both the first order and higher order types of uncertainty. Yet, the way uncertainty may be represented and processed also depends on the choice of what technique to use. There are different uncertainty representation and handling techniques that may be exploited in devising FRI mechanisms. It is therefore desirable to have a generic framework in which such techniques may be

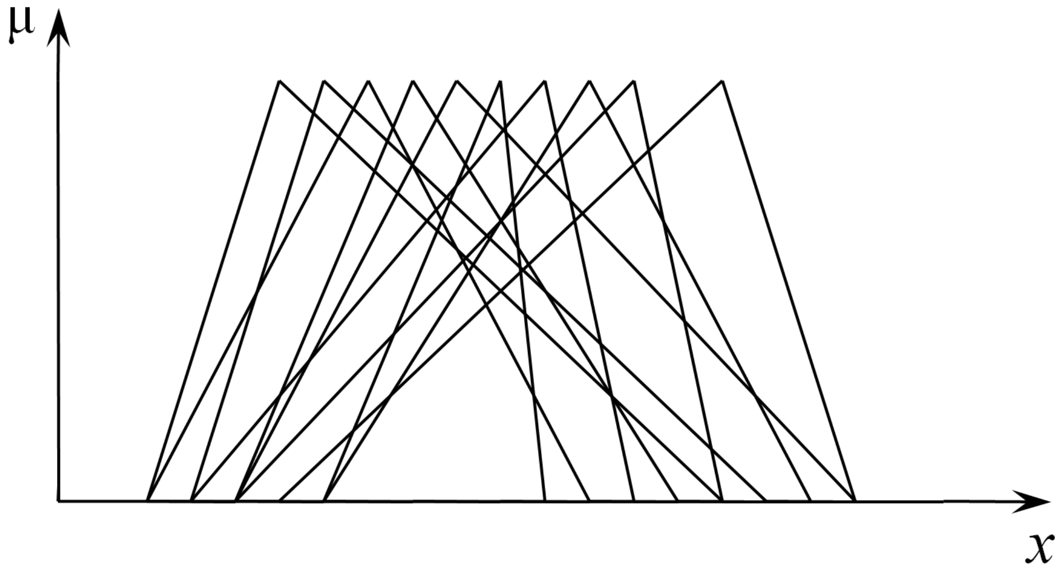


Figure 1.1: Different MFs for a common underlying concept perceived by different people

unified and further developed. For this reason, a novel framework is proposed in this thesis, for both representing the knowledge involving higher order uncertainty and facilitating interpolation with such knowledge.

The proposed framework is a generalisation of the transformation-based FRI techniques [76, 77], extending the applications of the existing mechanism to higher order environment. It consists of two main components: higher order knowledge representation and higher order rule interpolation. It aims to offer greater flexibility in handling different types of uncertainty that may be present in sparse rule bases and observations. Instead of addressing the first order uncertainty like conventional FRI methods, the proposed framework can handle both the first order and higher order uncertain information coherently. The work reflects the intuition that the more useful information is involved in the interpolation process, the better interpolated results may be obtained.

1.5 Thesis Structure

This section outlines the structure of the remainder of this thesis.

Chapter 2: Background

This chapter first presents an overview of the existing FRI approaches, and lays out the foundation of this project. In particular, two principal groups of FRI approaches, namely single step FRI and intermediate rule-based FRI, are reviewed, each being associated with a detailed description of a representative approach as well as its extensions and improvements. Then, basic knowledge representations for characterising different types of uncertainty are introduced, including rough sets (RSs) and type-2 FSs, which are each used to implement one version of the proposed framework. This chapter also describes the idea of information aggregation, which is the basis for the extension of the framework. The ordered weighted averaging (OWA) operators and the similarity measure operators are reviewed in detail.

Chapter 3: Framework for Higher Order Representation and Interpolation

This chapter proposes a novel transformation-based framework for both representing the knowledge involving higher order uncertainty and facilitating interpolation with such knowledge. It allows transformation-based rule interpolation techniques to be utilised in implementing a working fuzzy reasoning system. The framework can handle both the first order and higher order types of uncertainty coherently. The chapter presents the concept of higher order fuzzy sets (HOFSs) and the algorithm for higher order interpolation. The framework works by representing the knowledge involving uncertainty to higher order representation first and then, to derive the final conclusions by performing higher order interpolation over this representation.

Chapter 4: Implementing Framework with Rough-Fuzzy Sets

A rough-fuzzy (RF) implementation of the framework is presented in this chapter. Inspired by the concept of RSs, a specific definition of RF sets is proposed first in order to describe the range of uncertainty, which is characterised by the lower and upper approximation MFs. The proposed approach facilitates the representation of uncertain FS MFs with RF approximations, thereby improving the flexibility of rule interpolation in dealing with different types of uncertainty in fuzziness. An algorithm for RF rule interpolation is explained assuming that sparse rule bases involving RF-valued variables are available. It exploits the concept of RF sets and generalises the T-FRI techniques. This development has been published in [25, 26]. A proof of this generalisation is also provided.

Chapter 5: Implementing Framework with Type-2 Fuzzy Sets

This chapter describes another implementation of the framework using type-2 FSs and compares this alternative with the RF approach. For completeness, a comparison between type-2 FSs and RF sets is provided. The basic concepts involved are introduced and the implementation with type-2 FSs is described. As with the RF implementation, both interpolation and extrapolation involving multiple antecedent variables and multiple rules are provided. The experimental examples demonstrate that the proposed approach is of natural appeal for FRI while dealing with the uncertainty that conventional type-1 FRI techniques may otherwise be difficult to handle. The resultant mechanism is a useful extension of the existing type-1 FRI. The work developed in this chapter has been published in [27].

Chapter 6: Higher Order Fuzzy Rule Interpolation: Evaluations

In this chapter, the effectiveness of the proposed framework is illustrated by a practical application of predicting diarrhoeal disease rates in remote villages. Experts have always attempted to model how environmental change influences disease burden so that they can predict the disease rate. However, the models built for this are often very complicated and usually result in a sparse rule base. Moreover, different experts may have different kinds of expertise, resulting in similar but different expert rules and observations. Therefore, such problems provide a potentially suitable testbed for this framework. This application implies the potential of the framework in enhancing the robustness of FRI. Moreover, this framework is further evaluated in the application to other realistic decision making problems. The resultant accuracy reveals the efficacy of the framework.

Chapter 7: Theoretical Extension

This chapter extends the original definition of RF set-based FRI to a more general version, supported by the use of the OWA operators. The extended OWA-based FRI approach is then applied to group decision making (GDM) problems. The goal in GDM is to ensure that the best decision is made with respect to the available information and knowledge possessed by all group members. However, different types of uncertainty may influence both the assessment of the individual views and the derivation of the overall group-level solution. In the extended approach, individual preferences are firstly aggregated by means of a method derived from the use of RF set theory, and RF-based interpolation is then applied to derive the group-level conclusion. Experimental investigations are carried out and the results are presented to demonstrate the efficacy of the proposed work in guaranteeing the overall decision accuracy. The techniques described in this chapter have been published in [28].

Chapter 8: Conclusion

This chapter concludes the thesis with a summary of the achievements of the research presented, together with a discussion of possible future directions for research and potential areas for implementation of the work.

Appendices

Appendix A lists the publications arising from the work presented in this thesis, containing both published papers, and that currently under review for journal publication. Appendix B summaries the acronyms employed throughout this thesis.

Chapter 2

Background

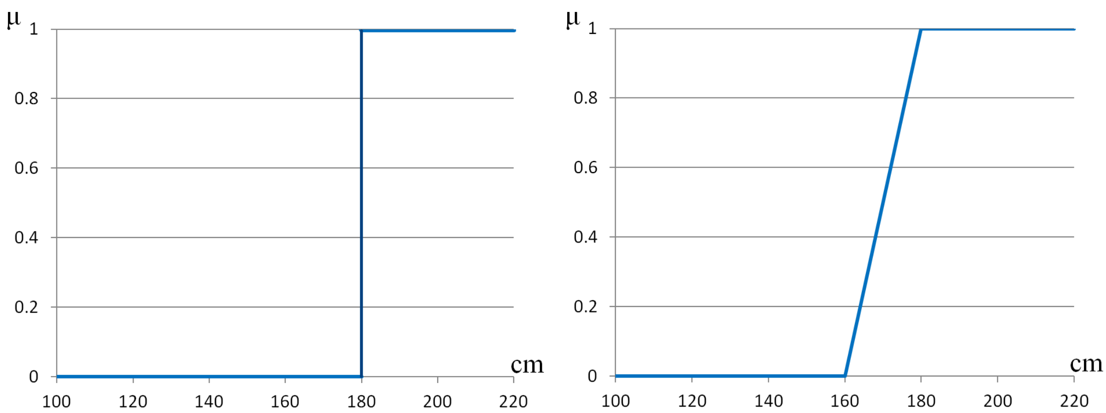
FUZZY rule interpolation (FRI) strengthens the power of fuzzy inference by enhancing the robustness of fuzzy inference systems (FISs) [76, 77, 206]. However, little existing work on FRI can conjunctively handle more than one form of uncertainty in the rules or observations. For instance, the difficulty in defining the required precise-valued membership functions (MFs) of the fuzzy sets (FSs) significantly restricts the application of conventional FRI techniques. To overcome such difficulties, this thesis presents significant developments in establishing novel FRI techniques. To set the background of these developments, this chapter reviews the relevant literature, including the existing FRI methods, rough sets (RSs), type-2 FSs and aggregation methods.

2.1 Fuzzy Set Theory

The modelling of imprecise and qualitative knowledge, as well as the transmission and handling of uncertainty at various stages are possible through the use of FSs [212]. Fuzzy logic is capable of supporting human type reasoning in natural form [132]. It is the earliest and most widely reported constituent of soft computing (SC). The development of fuzzy logic has led to the emergence of SC [142].

FSs are a further development of the mathematical concept of a set. An FS is an extension of a crisp set, where the latter allows only full membership or no membership at all, whereas the former allows partial membership. In a crisp set, membership or non-membership of an element is described by a characteristic function in the binary pair $\{0,1\}$. FS theory extends this concept by defining partial membership. An FS is characterised by a membership function (MF) that takes values in the interval $[0,1]$. In this case, a given element can be a member of more than one FS at a time.

As an example, consider the concept *tall*. In a crisp set, all of the people with height 180 cm or more are considered tall, and all of the people with height of less than 180 cm are considered not tall. The crisp set characteristic function is shown in Figure 2.1a, while the corresponding FS with a smooth MF is shown in Figure 2.1b, where X and Y axes denote the height and its corresponding membership value, respectively. The MF curve defines the transition from not tall and shows the degree of membership for any given height.



(a) Crisp characteristic function

(b) Fuzzy MF

Figure 2.1: Functions for height

Let X be the universe, an FS, A, in X is a set of ordered pairs

$$A = \{(x,\ \mu_A(x))|\mu_A(x) \in [0,1],\ x \in X\} \tag{2.1}$$

Such an FS is a collection of objects with graded membership, where $\mu_A(x)$ is termed the grade of membership of x in A. The closer the value of $\mu_A(x)$ is to 1, the more x belongs to the set A.

Essentially, an MF is a function that defines how each point in the input space is mapped to a membership value between 0 and 1. Various types of MFs can be used, including triangular, trapezoidal, Gaussian curves, polynomial curves, etc. In particular, due to the fact that triangular and trapezoidal FSs are commonly used in many FRI approaches [23, 73, 76, 77, 206, 208], they are therefore adopted for the work in this thesis. Other MFs (e.g., Gaussian) will be implemented in the future. Note that as using such continuous MFs, there will be no gap between any rules. In this case, however, FRI can still make sense above a certain minimum threshold in performing observation and rule matching. Triangular and trapezoidal MFs are defined respectively by three and four parameters and given by

$$f(x:a,b,c)=\begin{cases}0 & \text{if } x<a\\ \dfrac{x-a}{b-a} & \text{if } a\leq x\leq b\\ \dfrac{c-x}{c-b} & \text{if } b<x\leq c\\ 0 & \text{if } x>c\end{cases} \tag{2.2}$$

where a and c denote the left and right extreme points (with membership values of 0), and b denotes the normal point (with a membership value of 1).

$$f(x:a,b,c,d)=\begin{cases}0 & \text{if } x<a\\ \dfrac{x-a}{b-a} & \text{if } a\leq x<b\\ 1 & \text{if } b\leq x\leq c\\ \dfrac{d-x}{d-c} & \text{if } c<x\leq d\\ 0 & \text{if } x>d\end{cases} \tag{2.3}$$

where a and d denote the left and right extreme points (with membership values of 0), and b and c denote the normal points (with membership values of 1).

The *support* of an FS A is defined by

$$\text{supp}(A)=\{x\in X \mid \mu_A(x)>0\} \tag{2.4}$$

Its *core* is defined by

$$\text{core}(A)=\{x\in X \mid \mu_A(x)=1\} \tag{2.5}$$

An important property of FSs is their convexity. An FS A on X is *convex* if and only if [104]

$$\mu_A(\lambda x_1+(1-\lambda)x_2)\geq\min(\mu_A(x_1),\mu_A(x_2)) \tag{2.6}$$

for all $x_1,x_2\in X$ and all $\lambda\in[0,1]$.

An equivalent representation to the above standard definition is: an FS A is said to be *convex* if and only if

$$\mu_A(z) \geq \min(\mu_A(x), \mu_A(y)),\ \forall(x, y, z) \in X \text{ and } z \in [x, y] \tag{2.7}$$

where z is a point between x and y.

A is said to be *normal* if and only if

$$\mu_A(x) = 1,\ \exists x \in X \tag{2.8}$$

An arbitrary *polygonal* FS with n odd points, $A = (a_0, \ldots, a_{n-1})$, is shown in Figure 2.2. It has $\lfloor(n/2)\rfloor$ *supports* (horizontal intervals between every pair of odd points which have the same membership value) and $2(\lceil(n/2)\rceil - 1)$ *slopes* (non-horizontal intervals between every pair of consecutive odd points). In particular, two top points (of full membership value) do not have to be different.

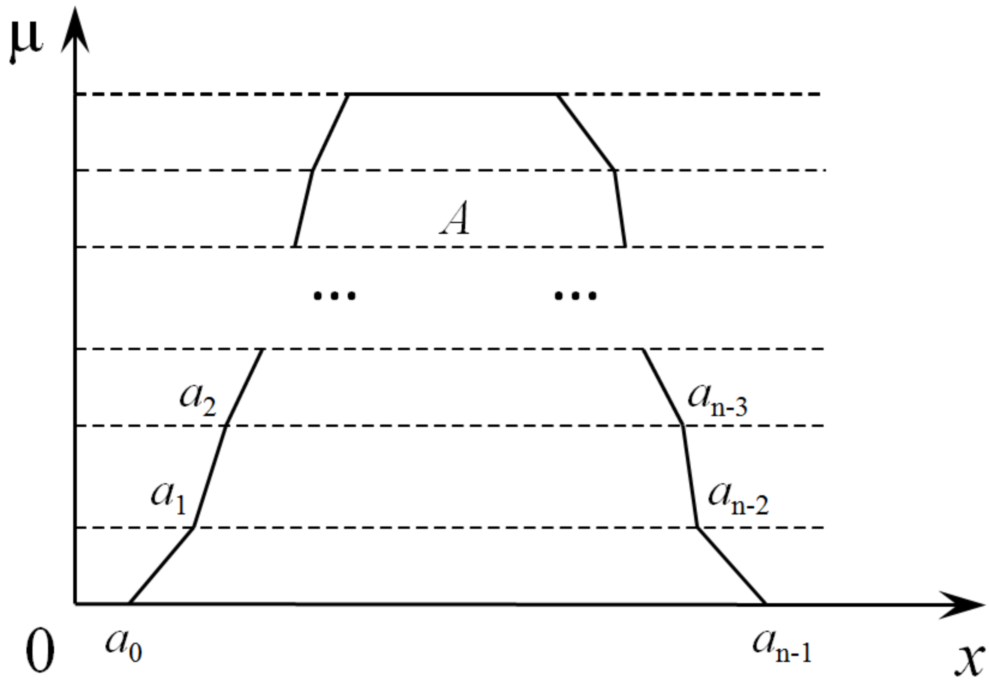


Figure 2.2: Polygonal FS

Note that although this figure explicitly assumes that evenly paired odd points are given at each α-cut level, this does not affect the generality of the FS representation as artificial odd points can be created to construct evenly paired odd points.

2.2 Structure of Fuzzy Inference

The process of fuzzy inference is basically an iteration of a computer paradigm based on FS theory, fuzzy IF-THEN rules and fuzzy reasoning. Each iteration takes inputs which can be an observation or a previously inferred crisp or fuzzy result. These inputs are then used to “fire” the rules in a given rule base. From this, the output is the aggregation of the inferred results from all of the fired rules. The general structure of fuzzy inference is illustrated in Figure 2.3.

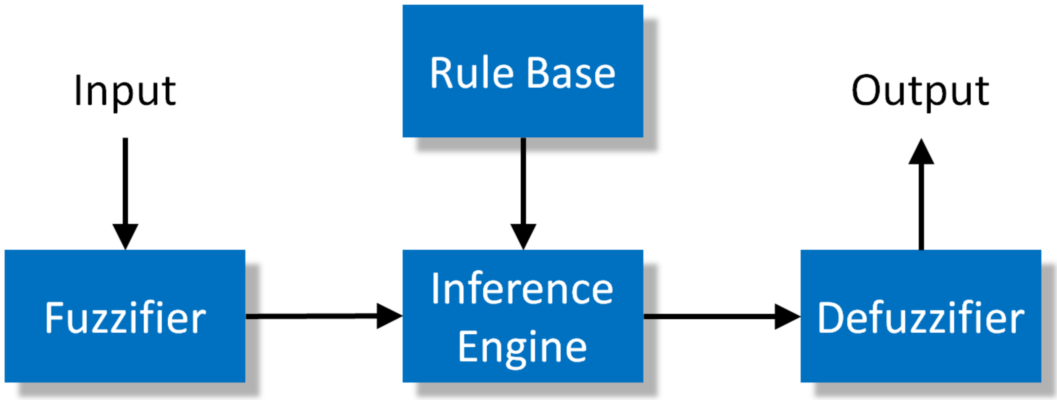


Figure 2.3: Generic FISs

- The *fuzzifier* maps discrete or real-valued inputs into corresponding fuzzy memberships. This is required in order to build rules that can be considered in terms of linguistic variables. The fuzzifier takes input values and determines the degree to which they belong to each of the FSs by means of MFs.
- The *rule base* contains linguistic rules that are provided by experts. It is also possible to extract rules from numerical data. Once the rules have been established, the FIS can be viewed as a system that maps an input vector to an output vector.
- The *inference engine* defines the mapping from input FSs into output FSs. It determines the degree to which the antecedent is satisfied for each rule. If the antecedent of a given rule has more than one part, fuzzy operators are applied to obtain a number that represents the result of the antecedents for that particular rule. Furthermore, if one or more rules fire simultaneously,

outputs for all rules are then aggregated. During the aggregation process, FSs that represent the output of each rule are combined into a single FS.

- The *defuzzifier* maps output FSs into a crisp or discrete output. Given an FS that encompasses a range of output values, the defuzzifier returns a single value. Several methods for defuzzification can be used in practice, including: centroid, maximum, etc.

2.3 Interpolative Reasoning Methods

Fuzzy systems use fuzzy rule bases to make inference. If the input domain is covered completely by the rule bases, such fuzzy rule bases are called *dense* rule bases [93]. In dense rule bases, for all the possible observations there exists at least one (at least partially) fired rule, whose antecedent part overlaps the input data. When an observation occurs, a consequence can be inferred by using conventional fuzzy reasoning methods such as Mamdani [129, 130] and TSK [167, 168]. On the contrary, for a *sparse* rule base, that is, the input domain is covered incompletely by the rule base, there is an empty space between two MFs of antecedents [164]. In this case, conventional fuzzy reasoning methods may encounter difficulty if an observation occurs in the empty space (which is also termed a “gap”), resulting in no rule fired and thus, no consequence derived. In general, the “empty space” is above a certain minimum confidence threshold if MFs like Gaussian are used.

The reasons for sparse or incomplete rule bases are various but have several aspects [183]:

- Originally, fuzzy systems were constructed from IF-THEN rules provided by human experts. More recently, learning techniques have increasingly been developed and applied to the construction of fuzzy IF-THEN rules from numerical data. However, both ways of constructing rule bases can result in sparse rule bases. In the former case, an incomplete rule base may be the consequence of missing expertise for certain system configurations. In the latter case, it may be that data used in the construction of the rule base does not sufficiently represent the input parameters.

- Fuzzy inference methods are often criticised when the number of inputs is large. The size of the rule base and the complexity of the inference algorithm grow exponentially with the number of inputs. A possible solution to reduce complexity is to omit redundant rules. This can, however, lead to incomplete rule bases [108].

- "Gaps" can be defined between rule bases intentionally, in order to avoid high complexity in large systems.

In the case where a fuzzy rule base contains "gaps", conventional fuzzy reasoning methods can no longer be used. This fact is due to the failure of traditional inference mechanisms in the case when observations find no fuzzy rule to fire. This cannot be allowed when using a fuzzy system in any practical application and such a system is considered useless. This problem was initially outlined in the "tomato classification" problem [143], shown in Figure 2.4.

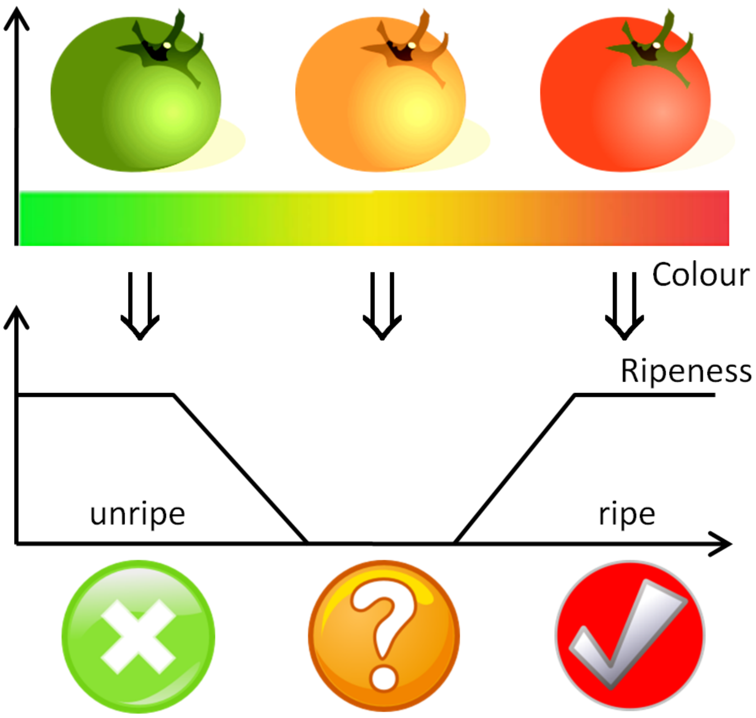


Figure 2.4: Fuzzy reasoning for the tomato problem

$$\begin{aligned}
\text{Rule 1} &: \text{If a tomato is red, then the tomato is ripe.} \\
\text{Rule 2} &: \text{If a tomato is green, then the tomato is upripe.} \\
\text{Observation} &: \text{This tomato is yellow.} \\
\text{Conclusion} &: \text{???}
\end{aligned}$$

The intuitive consequence of a human being would be that this tomato is half ripe. However, the MF "yellow" has no overlap with the MFs "red" or "green". Therefore, none of the conventional fuzzy inference mechanisms is able to reach such a conclusion.

Motivated by this, fuzzy interpolative reasoning mechanisms are proposed for performing fuzzy inference with systems comprising insufficient knowledge or sparse rule bases. Even when a given observation has no overlap with the antecedent values of any existing rules, FRI may still derive a useful conclusion. The techniques of FRI not only support inference in such situations, but also help to reduce the complexity of fuzzy models by eliminating the rules which may be approximated from their neighbouring rules.

A number of important FRI approaches have been proposed in the literature [90, 92, 156]. In terms of the underlying methodology, most of these approaches can be divided into two groups: single step rule interpolation and intermediate rule-based interpolation.

The first group of approaches directly interpolates a rule whose antecedent is identical to the given observation and thus, the consequence of the interpolated rule is the logical result of the observation. The most typical approach in this group is the first proposed FRI technique [106, 107], denoted the KH (Kóczy and Hirota) approach, which is based on the Decomposition Principle and Resolution Principle [104, 158, 222]. According to these principles, each FS can be represented by a series of α-cuts ($\alpha \in (0,1]$). Given α, the α-cut of the interpolated consequent FS can be calculated from the α-cuts of the (newly observed) antecedent FSs and all of the FSs involved in the rules used for interpolation. Having found the α-cuts of the consequent FS for all $\alpha \in (0,1]$, the consequent FS is then assembled by applying the Resolution Principle.

The second group of approaches reaches the target in two steps. In the first step these approaches interpolate an artificial intermediate rule. The antecedent of this intermediate rule is expected to be very close to the given observation. As a result, the interpolation problem becomes similarity reasoning [56, 103, 210]. The estimated conclusion is then derived in the second step according to the similarity between the observation and the antecedent of the artificial intermediate rule. The scale and move transformation-based FRI approach (T-FRI) [76, 77], which has been adopted as the foundation for the work in this thesis, belongs to this group.

As the two representatives for these two groups, the KH and T-FRI approaches are respectively reviewed in the following sections.

2.3.1 The KH Approach

The KH approach [106, 107] determines the conclusion by its α-cuts in such a way that the proportional distance between the estimated conclusion and the consequent sets of the rules which are used should be the same as the distance between the observation and the antecedents of those rules, for all important α-cuts. The x-cut A_α of an FS A is a crisp set, denoted: $A_\alpha = \{x|A(x) \geq \alpha, \alpha \in (0,1]\}$.

2.3.1.1 Base Case of the KH Approach

The starting ideas are the Extension Principle and Resolution Principle. The former states that the solution of a problem for FSs can be found in the form of solving first for arbitrary α-cuts that are crisp sets and then extending the solution to the fuzzy case. The latter describes the decomposition of FSs to α-cuts

$$\mu_A(x) = \sup\{\alpha : x \in A_\alpha\} \tag{2.9}$$

Every FS can be approximated with the use of the family of its α-cuts. Theoretically, all infinite cuts should be treated separately. In most practical cases, however, if the MF is piecewise linear, it is often sufficient to calculate its α-cuts for only a few important or typical values [164], e.g., $\alpha = 0$ and $\alpha = 1$.

An important concept in the KH approach is the "less than" relation between two convex and normal FSs. FS A_1 is said to be *less than* FS A_2, denoted by $A_1 \prec A_2$, if $\forall \alpha \in (0,1]$, the following conditions hold:

$$\inf\{A_{1\alpha}\} < \inf\{A_{2\alpha}\},\ \sup\{A_{1\alpha}\} < \sup\{A_{2\alpha}\} \tag{2.10}$$

where $A_{1\alpha}$ and $A_{2\alpha}$ are the α-cut sets of A_1 and A_2, respectively, $\inf\{A_{i\alpha}\}$ is the infimum of $A_{i\alpha}$, and $\sup\{A_{i\alpha}\}$ is the supremum of $A_{i\alpha}$, $i = 1,2$.

For simplicity, suppose that two single-antecedent fuzzy rules are given as follows:

$$R_1 : \text{If } x \text{ is } A_1, \text{ then } y \text{ is } B_1$$
$$R_2 : \text{If } x \text{ is } A_2, \text{ then } y \text{ is } B_2$$

They are said to be *neighbouring rules* if and only if: (1) $A_1 \prec A_2$ or $A_2 \prec A_1$; and (2) there is no individual rule "If x is A', then y is B'" such that $A_1 \prec A' \prec A_2$ if $A_1 \prec A_2$, or $A_2 \prec A' \prec A_1$ if $A_2 \prec A_1$.

To implement interpolation in the region between the antecedents of these two rules, i.e., to generate an approximated conclusion when an observation A^* located between FSs A_1 and A_2 is hereby given. The neighbouring rules in a given rule base are therefore said to flank the observation. For the above two rules, this means that $A_1 \prec A^* \prec A_2$ or $A_2 \prec A^* \prec A_1$.

The KH approach uses the following equation to determine the interpolated result:

$$\frac{d(A^*,A_1)}{d(A^*,A_2)} = \frac{d(B^*,B_1)}{d(B^*,B_2)} \tag{2.11}$$

where A_1, A_2 are the antecedents of the two flanking rules, A^* is a given observation, B_1, B_2 are the consequences of those rules, B^* is the estimated conclusion, and $d(.,.)$ is typically the Euclidean distance between two FSs (though other distance metrics may be also used).

According to the Decomposition Principle, a convex and normal FS A can be represented by a series of α-cut intervals, each denoted as A_α, $\alpha \in (0,1]$. In this case, Equation (2.11) can be written as:

$$\frac{d(A^*_\alpha,A_{1\alpha})}{d(A^*_\alpha,A_{2\alpha})} = \frac{d(B^*_\alpha,B_{1\alpha})}{d(B^*_\alpha,B_{2\alpha})} \tag{2.12}$$

where given any α ($\alpha \in (0,1]$), the lower and upper distances between α-cuts $A_{1\alpha}$ and $A_{2\alpha}$ are defined:

$$\begin{cases} d_L(A_{1\alpha},A_{2\alpha}) = d(\inf\{A_{1\alpha}\}, \inf\{A_{2\alpha}\}) \\ d_U(A_{1\alpha},A_{2\alpha}) = d(\sup\{A_{1\alpha}\}, \sup\{A_{2\alpha}\}) \end{cases} \tag{2.13}$$

Note that the Euclidean distance between intervals can be defined in different ways but they all lie between $d_L(A_{1\alpha}, A_{2\alpha})$ and $d_U(A_{1\alpha}, A_{2\alpha})$. From Equation (2.13), Equation (2.12) can be rewritten as

$$\begin{aligned}\frac{d_L(A^*_\alpha, A_{1\alpha})}{d_L(A^*_\alpha, A_{2\alpha})} &= \frac{d_L(B^*_\alpha, B_{1\alpha})}{d_L(B^*_\alpha, B_{2\alpha})} \\ &= \frac{d_L(\inf\{B^*_\alpha\}, \inf\{B_{1\alpha}\})}{d_L(\inf\{B^*_\alpha\}, \inf\{B_{2\alpha}\})} \\ &= \frac{\inf\{B^*_\alpha\} - \inf\{B_{1\alpha}\}}{\inf\{B_{2\alpha}\} - \inf\{B^*_\alpha\}}\end{aligned} \tag{2.14}$$

Equation (2.14) can then be solved as follows:

$$\begin{aligned}\inf\{B^*_\alpha\} &= \frac{\inf\{B_{1\alpha}\} d_L(A^*_\alpha, A_{2\alpha}) + \inf\{B_{2\alpha}\} d_L(A^*_\alpha, A_{1\alpha})}{d_L(A^*_\alpha, A_{2\alpha}) + d_L(A^*_\alpha, A_{1\alpha})} \\ &= \frac{\frac{\inf\{B_{1\alpha}\}}{d_L(A^*_\alpha, A_{1\alpha})} + \frac{\inf\{B_{2\alpha}\}}{d_L(A^*_\alpha, A_{2\alpha})}}{\frac{1}{d_L(A^*_\alpha, A_{1\alpha})} + \frac{1}{d_L(A^*_\alpha, A_{2\alpha})}}\end{aligned} \tag{2.15}$$

where $\sup\{B^*_\alpha\}$ can be calculated in the same way, resulting in

$$\begin{cases} \inf\{B^*_\alpha\} = \dfrac{\frac{\inf\{B_{1\alpha}\}}{d_L(A^*_\alpha, A_{1\alpha})} + \frac{\inf\{B_{2\alpha}\}}{d_L(A^*_\alpha, A_{2\alpha})}}{\frac{1}{d_L(A^*_\alpha, A_{1\alpha})} + \frac{1}{d_L(A^*_\alpha, A_{2\alpha})}} \\ \sup\{B^*_\alpha\} = \dfrac{\frac{\sup\{B_{1\alpha}\}}{d_U(A^*_\alpha, A_{1\alpha})} + \frac{\sup\{B_{2\alpha}\}}{d_U(A^*_\alpha, A_{2\alpha})}}{\frac{1}{d_U(A^*_\alpha, A_{1\alpha})} + \frac{1}{d_U(A^*_\alpha, A_{2\alpha})}} \end{cases} \tag{2.16}$$

Alternatively, let

$$\begin{cases} \lambda_L = \dfrac{d_L(A^*_\alpha, A_{1\alpha})}{d_L(A^*_{2\alpha}, A_{1\alpha})} \\ \lambda_U = \dfrac{d_U(A^*_\alpha, A_{1\alpha})}{d_U(A^*_{2\alpha}, A_{1\alpha})} \end{cases} \tag{2.17}$$

The same solution can then be obtained but represented differently as follows:

$$\begin{cases} \inf\{B^*_\alpha\} = (1-\lambda_L)\inf\{B_{1\alpha}\} + \lambda_L \inf\{B_{2\alpha}\} \\ \sup\{B^*_\alpha\} = (1-\lambda_U)\sup\{B_{1\alpha}\} + \lambda_U \sup\{B_{2\alpha}\} \end{cases} \tag{2.18}$$

From this, $B^*_\alpha = [\inf\{B^*_\alpha\}, \sup\{B^*_\alpha\}]$ results. The estimated conclusion B^* can then be constructed by using the representation principle of FSs:

$$B^* = \bigcup_{\alpha \in (0,1]} \alpha B^*_\alpha \tag{2.19}$$

The most important advantage of the KH approach is its low computational complexity that ensures the fast response performance for real-time applications. Despite the rapid development of α-cut based FRI, there is a drawback in this group of methods. Theoretically, all possible α-cuts (an infinite number) should be considered in performing the interpolation. However, the existing approaches in this group only take a finite number of α-cuts into consideration (usually 3 or 4). The resulting points are then connected by linear pieces to produce an approximation of the accurate conclusion.

2.3.1.2 Extensions of the KH Approach

The principle of interpolating two rules can be extended in many different ways. A possible way to generalise the KH approach is to increase the number of the involved fuzzy rules that are taken into consideration during the computation of the conclusion.

Suppose that N fuzzy rules flank the observation from both sides in the sense of $\prec$. Intuitively, the further a given fuzzy rule is located from the observation, the less weight the respective consequence in the construction of the conclusion play. This can be obtained from the solution of Equation (2.16) repeatedly for the pairs of points and by averaging the various solutions in a weighted way. The overall solutions are as follows:

$$\begin{cases} \min\{B^*_\alpha\} = \dfrac{\sum\limits_{i=1}^{2N} \frac{1}{d_L(A^*_\alpha, A_{i\alpha})} \inf\{B_{i\alpha}\}}{\sum\limits_{i=1}^{2N} \frac{1}{d_L(A^*_\alpha, A_{i\alpha})}} \\ \max\{B^*_\alpha\} = \dfrac{\sum\limits_{i=1}^{2N} \frac{1}{d_U(A^*_\alpha, A_{i\alpha})} \sup\{B_{i\alpha}\}}{\sum\limits_{i=1}^{2N} \frac{1}{d_U(A^*_\alpha, A_{i\alpha})}} \end{cases} \tag{2.20}$$

2.3.1.3 Modifications of the KH Approach

One disadvantage of the KH approach is that the membership of the derived FS is not always a function leaving alone to be a fuzzy MF, which is shown in Figure 2.5. The recognition of the "abnormal problem" of the KH approach has led to the development of many techniques, which modify or improve the original approach.

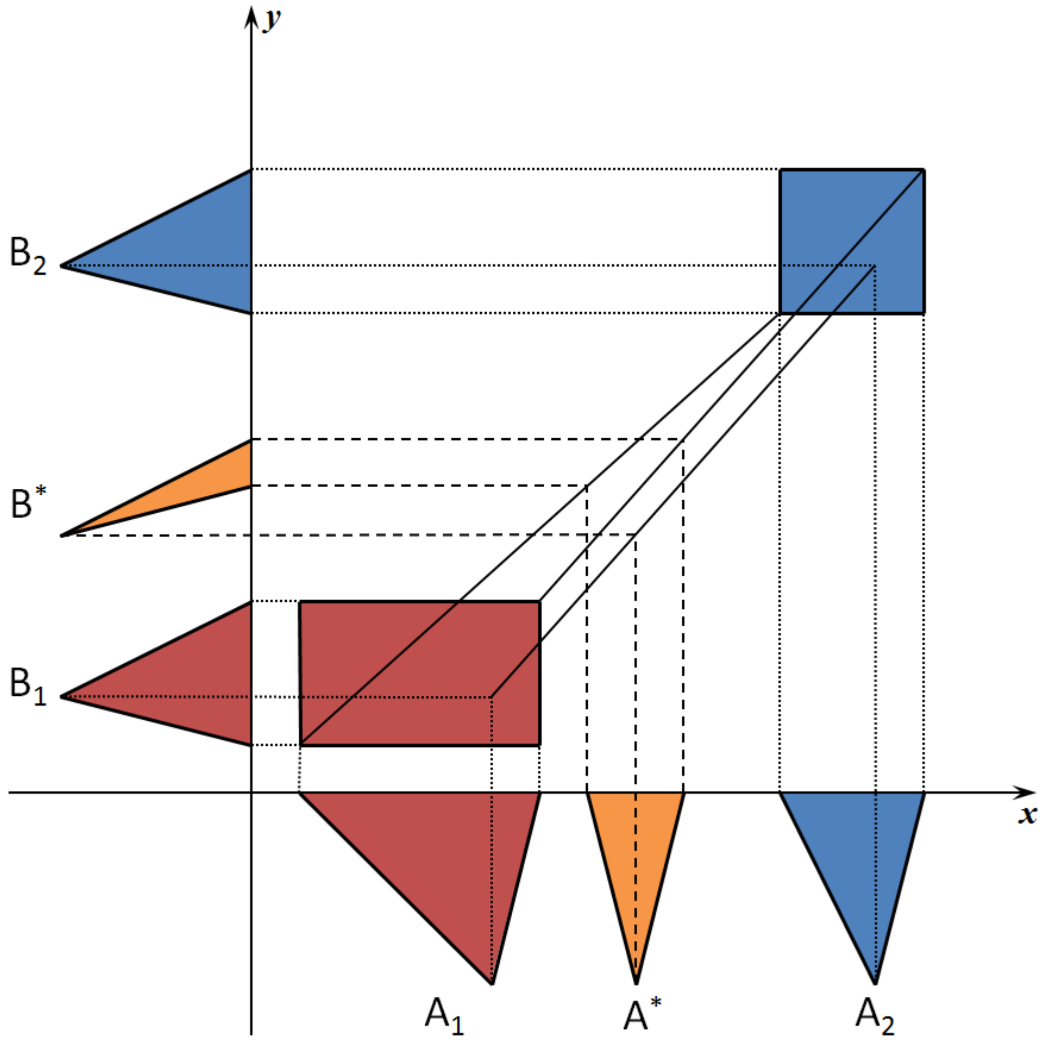


Figure 2.5: Abnormal conclusion generated by the KH approach

The VKK (Vass, Kalmár, and Kóczy) approach [176] modifies the distance measure defined in the KH approach. It describes each α-cut by its centre point and its width. The distance between two FSs is characterised by a vector which contains a set of distances between each corresponding pair of α-cuts of the two FSs. This approach is also applicable for interpolation with multiple antecedent rules, which is achieved by aggregating the distances on different antecedent attributes of a certain level by Euclidean distance and calculating the resultant width using the arithmetic average. However, this approach is not applicable for problems with singleton observations because the α-cut width of 0 is not considered.

The modified α-cut based interpolation (MACI) [169] solves the abnormality problem effectively, while it maintains the advantageous properties of the KH approach itself. This approach represents each FS with two vectors which describe the left (lower) and the right (upper) flank by means of the technique published in [203]. The vectors contain the break points in case of piecewise linear MFs or endpoints of predefined α-cuts in case of smooth MFs. However, this approach also does not preserve linearity, but the deviation of the piecewise linear conclusion from the accurate one is less than in the case of the original approach.

2.3.2 The T-FRI Approach

The T-FRI approach [76, 77] can handle both interpolation and extrapolation of multiple multi-antecedent rules with triangular, complex polygon, Gaussian and bell-shaped fuzzy MFs. It has the following properties:

- It can handle both interpolation and extrapolation which involve multiple fuzzy rules, with each rule consisting of multiple antecedents.
- It guarantees the uniqueness as well as normality and convexity of the resulting interpolated FSs.
- It preserves piece-wise linearity such that interpolation can be computed using only characteristic points which describe a given polygonal FS, thereby ignoring any non-characteristic points and saving computation effort.
- It has been applied to problems such as truck backer-upper control and computer activity prediction.

2.3.2.1 Representative Value

A key concept used in the T-FRI approach is the *representative value* (Rep) of a given FS, it captures important information such as the overall location of an FS.

Consider an arbitrary polygonal FS A with k odd points, which can be denoted as $A = (a_0, \cdots, a_{k-1})$. Given such an arbitrary polygonal FSs, its general Rep is defined by

$$\mathrm{Rep}(A) = \sum_{i=0}^{k-1} w_i a_i \tag{2.21}$$

where w_i is the weight assigned to the point a_i.

In general, the specification of the weights is necessary for a given application. Different definitions can be adopted for deriving different Rep values. The simplest case is that all points take the same weighting value, i.e.,

$$\mathrm{Rep}(A) = \frac{1}{k}\sum_{i=0}^{k-1} a_i \tag{2.22}$$

An alternative is the *weighted average Rep*, where the weights increase upwards from the bottom support to the top support, to reflect the relative significance of the fuzzy membership values. For instance, assuming the weights increase upwards from 0.5 to 1, such a Rep is defined by

$$\mathrm{Rep}(A) = \frac{\sum_{i=0}^{\lceil \frac{k}{2} \rceil - 1} \frac{1+\mu_i}{2}(a_i + a_{k-i-1})}{\sum_{i=0}^{\lceil \frac{k}{2} \rceil - 1} \frac{1+\mu_i}{2}} \tag{2.23}$$

where μ_i is the membership value of a_i.

Note that artificial odd points can be created to construct evenly paired odd points (as indicated previously), so $\mu_i = \mu_{k-i-1}$ can always be assumed.

One of the most widely used defuzzification methods, the centre of core, can also be utilised as an alternative. The *centre of core Rep* is solely determined by those points with a fuzzy membership value of 1:

$$\mathrm{Rep}(A) = \frac{1}{2}(a_{\lceil \frac{k}{2} \rceil - 1} + a_{k - \lceil \frac{k}{2} \rceil}) \tag{2.24}$$

Based on the generated Rep values, the interpolation process is discussed in the following three cases. For simplicity, only rules involving triangular-shaped MFs are considered.

2.3.2.2 The T-FRI Approach with Two Single-antecedent Rules

Suppose that two neighbouring rules $A_1 \Rightarrow B_1, A_2 \Rightarrow B_2$ and an observation A^*, which is located between FSs A_1 and A_2, are given as follows:

$$\begin{aligned}
R_1 &: \text{if } x_1 \text{ is } A_1, \text{ then } y_1 \text{ is } B_1 \\
R_2 &: \text{if } x_2 \text{ is } A_2, \text{ then } y_2 \text{ is } B_2 \\
O &: x \text{ is } A^* \\
C &: y \text{ is } B^*
\end{aligned}$$

The desired conclusion B^* can be derived by interpolation. An intermediate rule $A' \Rightarrow B'$ is first constructed by manipulating these two given rules, where the intermediate term A' and the observation A^* have the same Rep, and so do the intermediate term B' and the desired B^*. Then B' is converted into B^* using scale and move transformations, which have been used to transform A' to A^*.

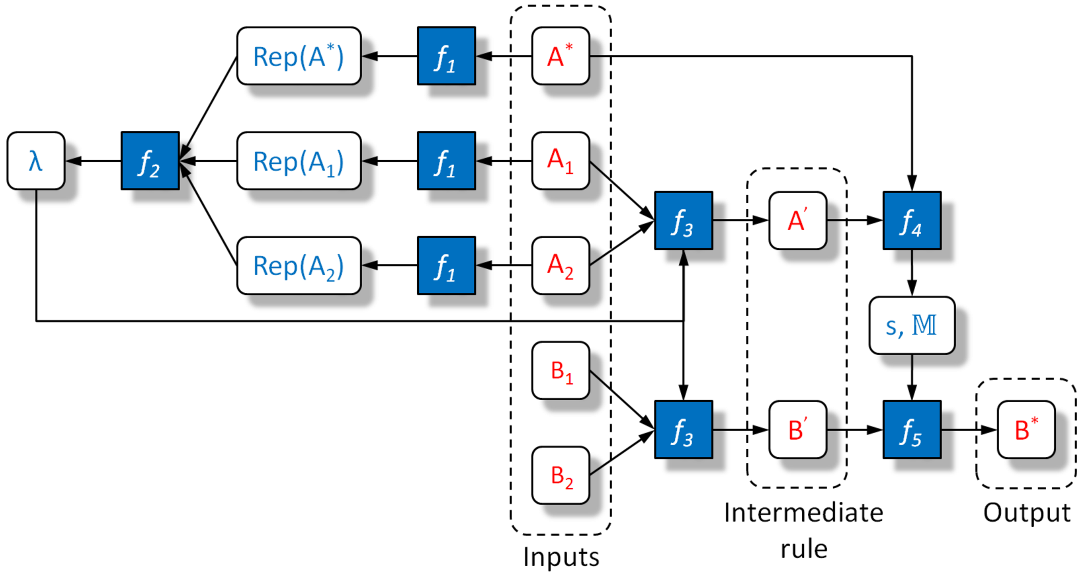


Figure 2.6: T-FRI with two single-antecedent rules

The interpolation process is illustrated in Figure 2.6. Given FSs A^*, A_1 and A_2, three parameters Rep(A^*), Rep(A_1) and Rep(A_2) are produced with the function f_1. Next, the relative placement relation between the observation A^* and the antecedents (A_1 and A_2) of the two neighbouring rules is calculated by the function f_2, resulting in λ. From this, an intermediate rule $A' \Rightarrow B'$ is generated by applying the function f_3 with parameter λ to both the antecedents and consequences of the neighbouring rules. Then, the similarity degree between A' and A^* is computed by a predefined similarity measure. Specifically, scale rate s and move ratio $\mathbb{M}$ are exploited in scale and move transformation-based interpolation to represent the similarity degree, which is achieved by the function f_4. Finally, the estimated conclusion B^* is obtained by applying the function f_5 to B' while imposing the same similarity degree.

Intermediate Rule

The *relative placement factor* λ of the observation A^*, with respect to its two neighbouring rule antecedents A_1 and A_2, is defined by

$$\begin{aligned}\lambda &= \frac{d(A_1,A^*)}{d(A_1,A_2)} \\ &= \frac{d(\mathrm{Rep}(A_1),\mathrm{Rep}(A^*))}{d(\mathrm{Rep}(A_1),\mathrm{Rep}(A_2))}\end{aligned} \tag{2.25}$$

where $d(A_1,A_2) = d(\mathrm{Rep}(A_1),\mathrm{Rep}(A_2))$ represents the distance between two FSs A_1 and A_2, which is defined by

$$\begin{aligned}d(A_1,A_2) &= d(\mathrm{Rep}(A_1),\mathrm{Rep}(A_2)) \\ &= \mathrm{Rep}(A_2) - \mathrm{Rep}(A_1)\end{aligned} \tag{2.26}$$

where $\mathrm{Rep}(A_1) \neq \mathrm{Rep}(A_2)$ because $A_1 \prec A_2$ or $A_2 \prec A_1$. Such a factor reflects the relative location of the interpolated rule regarding the two neighbouring rules.

By using the simplest linear interpolation, the antecedent of the intermediate rule $A' = (a'_0, a'_1, a'_2)$ can be calculated as follows:

$$\begin{cases} a'_0 = (1-\lambda)a_{10} + \lambda a_{20} \\ a'_1 = (1-\lambda)a_{11} + \lambda a_{21} \\ a'_2 = (1-\lambda)a_{12} + \lambda a_{22} \end{cases} \tag{2.27}$$

which are collectively abbreviated to

$$A' = (1-\lambda)A_1 + \lambda A_2 \tag{2.28}$$

In so doing, the Rep of the calculated A' is guaranteed to be equal to that of A^*. The consequence of the intermediate rule $B' = (b'_0, b'_1, b'_2)$ can then be obtained similar to the calculation of A' using the same λ:

$$\begin{cases} b'_0 = (1-\lambda)b_{10} + \lambda b_{20} \\ b'_1 = (1-\lambda)b_{11} + \lambda b_{21} \\ b'_2 = (1-\lambda)b_{12} + \lambda b_{22} \end{cases} \tag{2.29}$$

with abbreviated notation

$$B' = (1-\lambda)B_1 + \lambda B_2 \tag{2.30}$$

Scale and Move Transformations

As $A' \Rightarrow B'$ is derived from $A_1 \Rightarrow B_1$ and $A_2 \Rightarrow B_2$, it is feasible to perform fuzzy reasoning with this new rule without further reference to its originals. Given such an intermediate rule and an observation, the conclusion can be calculated with respect to the following intuition:

The more similar A' to A^*, the more similar B' to B^*.

Suppose that a certain degree of similarity between the antecedent parts A' and A^* is established, it is intuitive to require that the consequent parts B' and B^* attain the same similarity degree. Hence, the following two transformations are used to ensure this.

Scale Transformation The similarity degree between A' and A^* is first measured by *scale rate* s, which is defined by

$$s = \frac{a_2^* - a_0^*}{a_2' - a_0'} \tag{2.31}$$

Let $A'' = (a_0'', a_1'', a_2'')$ denote the second intermediate term generated by the scale transformation. By using s, the current support (a_0', a_2') is transformed into a new support (a_0'', a_2''), while keeping the Rep and the ratio of the left-support (a_0'', a_1'') to the right-support (a_1'', a_2'') of A'' the same as those of its original, such that

$$\begin{cases} a_2'' - a_0'' = s(a_2' - a_0') \\ \dfrac{a_0'' + a_1'' + a_2''}{3} = \dfrac{a_0' + a_1' + a_2'}{3} \\ \dfrac{a_1'' - a_0''}{a_2'' - a_1''} = \dfrac{a_1' - a_0'}{a_2' - a_1'} \end{cases} \tag{2.32}$$

A'' can then be calculated by solving Equation (2.32):

$$\begin{cases} a_0'' = \dfrac{a_0'(1+2s) + a_1'(1-s) + a_2'(1-s)}{3} \\ a_1'' = \dfrac{a_0'(1-s) + a_1'(1+2s) + a_2'(1-s)}{3} \\ a_2'' = \dfrac{a_0'(1-s) + a_1'(1-s) + a_2'(1+2s)}{3} \end{cases} \tag{2.33}$$

This measure reflects the similarity degree between A' and A^*: the closer is s to 1, the more similar is A' to A^*. It is therefore used to act as, or to contribute to, the desirable similarity degree in order to transform B' to B^*.

Move Transformation The similarity degree is further measured by *move ratio* $\mathbb{M}$. By using $\mathbb{M}$, the current support (a''_0, a''_2) of A'' is moved to (a^*_0, a^*_2) while keeping its Rep, resulting in the observation A^*. The move ratio $\mathbb{M}$ is defined by

$$\mathbb{M} = \begin{cases} \dfrac{a^*_0 - a''_0}{\frac{a''_1 - a''_0}{3}} & \text{if } a^*_0 \geq a''_0 \\ \dfrac{a^*_0 - a''_0}{\frac{a''_2 - a''_1}{3}} & \text{otherwise} \end{cases} \tag{2.34}$$

Given $\mathbb{M}$, A^* can then be retrieved as:

$$\begin{cases} \begin{cases} a^*_0 = a''_0 + \mathbb{M}\frac{a''_1 - a''_0}{3} \\ a^*_1 = a''_1 - 2\mathbb{M}\frac{a''_1 - a''_0}{3} \\ a^*_2 = a''_2 + \mathbb{M}\frac{a''_1 - a''_0}{3} \end{cases} & \text{if } \mathbb{M} \geq 0 \qquad (2.35a) \\ \begin{cases} a^*_0 = a''_0 + \mathbb{M}\frac{a''_2 - a''_1}{3} \\ a^*_1 = a''_1 - 2\mathbb{M}\frac{a''_2 - a''_1}{3} \\ a^*_2 = a''_2 + \mathbb{M}\frac{a''_2 - a''_1}{3} \end{cases} & \text{otherwise} \qquad (2.35b) \end{cases}$$

This reflects the similarity degree between A' and A^*: the closer is $\mathbb{M}$ to 0, the more similar is A' to A^*.

Having obtained the similarity degree between A' and A^*, the interpolated conclusion B^* can therefore be obtained by transforming B' with the same scale rate s and move ratio $\mathbb{M}$.

General Scale and Move Transformations

The general scale and move transformations for polygonal FSs can be extended from the previous subsection.

General Scale Transformation Consider A' and A^*, respectively represented as $A' = (a'_0, \cdots, a'_{k-1})$ and $A^* = (a^*_0, \cdots, a^*_{k-1})$. The following parameters, termed the general scale rates s_p $(p = 0, \cdots, \lfloor (k/2) \rfloor - 1)$ rescale the pth support of A' to approximate that of A^*:

$$s_p = \frac{a^*_{k-p-1} - a^*_p}{a'_{k-p-1} - a'_p} \tag{2.36}$$

From these general scale rates, the following general scale ratios $\mathbb{S}_q$ $(q = 1, \cdots, \lfloor (k/2) \rfloor - 1)$ modify the rescaled qth support of A' to further approximate that of A^* such that the resulting FS A' is of the same scale as that of A^*:

$$\mathbb{S}_q = \begin{cases} \dfrac{\frac{a^*_{k-q-1}-a^*_q}{a^*_{k-q}-a^*_{q-1}} - \frac{a'_{k-q-1}-a'_q}{a'_{k-q}-a'_{q-1}}}{1-\frac{a'_{k-q-1}-a'_q}{a'_{k-q}-a'_{q-1}}} & \text{if } s_q \geq s_{q-1} \\ \dfrac{\frac{a^*_{k-q-1}-a^*_q}{a^*_{k-q}-a^*_{q-1}} - \frac{a'_{k-q-1}-a'_q}{a'_{k-q}-a'_{q-1}}}{\frac{a'_{k-q-1}-a'_q}{a'_{k-q}-a'_{q-1}}} & \text{if } s_{q-1} > s_q \end{cases} \tag{2.37}$$

From this, by imposing the required similarities, the corresponding general scale rates s'_p that will help rescale the pth support of $\tilde{B}'$ into the emerging $\tilde{B}^*$ can be obtained such that

$$s'_p = \begin{cases} s_p & \text{if } p = 0 \\ \dfrac{s'_{p-1}(s_p - s_{p-1})(\frac{\tilde{b}'_{k-p}-\tilde{b}'_{p-1}}{\tilde{b}'_{k-p-1}-\tilde{b}'_p} - 1)}{s_{p-1}(\frac{a'_{k-p}-a'_{p-1}}{a'_{k-p-1}-a'_p} - 1)} + s'_{p-1} & \text{if } s_p \geq s_{p-1},\ p > 0 \\ \dfrac{s'_{p-1}s_p}{s_{p-1}} & \text{if } s_{p-1} > s_p,\ p > 0 \end{cases} \tag{2.38}$$

General Move Transformation The general move ratios $\mathbb{M}_r$ $(r = 0, \cdots, \lceil (k/2) \rceil - 2)$ shift the locations of supports of $A^{(r-1)}$ to that of A^* (where $A^{(r-1)}$ is the term obtained after the (r-1)th sub-move with initialisation $A^{-1} = A''$):

$$\mathbb{M}_r = \begin{cases} \dfrac{a^*_r - a^{(r-1)}_r}{\min\{\frac{a^{(r-1)}_r+\cdots+a^{(r-1)}_{\lceil (k/2) \rceil -1}}{\lceil (k/2) \rceil - r} - a^{(r-1)}_r, a^{(r-1)}_{k-r} - a^{(r-1)}_{k-r-1}\}} & \text{if } a^*_r \geq a^{(r-1)}_r \\ \dfrac{a^*_r - a^{(r-1)}_r}{\min\{a^{(r-1)}_{k-r-1} - \frac{a^{(r-1)}_{k-\lceil (k/2) \rceil}+\cdots+a^{(r-1)}_{k-r-1}}{\lceil (k/2) \rceil - r}, a^{(r-1)}_r - a^{(r-1)}_{r-1}\}} & \text{if } a^{(r-1)}_r > a^*_r \end{cases} \tag{2.39}$$

where $a^{(r-1)}_r$ is the a''_r's new position after the (r-1)th sub-move. Initially, when $r = 0$, $a^{(-1)}_0 = a''_0$, $a^{(r-1)}_{k-r} - a^{(r-1)}_{k-r-1}$ and $a^{(r-1)}_r - a^{(r-1)}_{r-1}$ are not included into the calculation of min{.,.}.

2.3.2.3 The T-FRI Approach with Two Multi-antecedent Rules

Two multi-antecedent rules interpolation is a generalisation of the two single-antecedent rules interpolation. Given an observation such that

$$O: \; x_1 \text{ is } A_1^*, \cdots, x_j \text{ is } A_j^*, \cdots, x_M \text{ is } A_M^*$$

Suppose that two neighbouring rules are used for interpolation with respect to the given observation, which are represented by

$$R_1: \text{If } x_1 \text{ is } A_{11}, \cdots, x_j \text{ is } A_{1j}, \cdots, x_M \text{ is } A_{1M}, \text{ then } y \text{ is } B_1$$
$$R_2: \text{If } x_1 \text{ is } A_{21}, \cdots, x_j \text{ is } A_{2j}, \cdots, x_M \text{ is } A_{2M}, \text{ then } y \text{ is } B_2$$

where M is the number of antecedent variables.

When one rule involves multiple antecedent variables, each antecedent dimension will have its own parameter values for λ, s and $\mathbb{M}$. Obviously, all these values contribute to the construction of the intermediate term B' and the desired B^*. The following equations aggregate all of these values in order to construct the intermediate term B'. The interpolated conclusion B^* can then be obtained by using s' and $\mathbb{M}'$, where

$$\lambda' = \frac{1}{M}\sum_{j=1}^{M}\lambda_j \tag{2.40}$$

$$B' = (1-\lambda')B_1 + \lambda' B_2 \tag{2.41}$$

$$s' = \frac{1}{M}\sum_{j=1}^{M}s_j \tag{2.42}$$

$$\mathbb{M}' = \frac{1}{M}\sum_{j=1}^{M}\mathbb{M}_j \tag{2.43}$$

and M is the number of antecedent variables.

The process of the T-FRI with two multi-antecedent rules is illustrated in Figure 2.7. In this figure, there are M repeated components which are identical to the core of the two single-antecedent rules interpolation (as shown in Figure 2.6). Each of these components does exactly the same as the common core of the single-antecedent situation. That is, the relative placement factors λ_j $(j = 1, \ldots, M)$ are calculated from each term of the observation A_j^* and the corresponding two FSs A_{1j} and A_{2j}. The function f_6 is then introduced to combine all these λ_j to a single parameter λ', resulting in the consequence of the intermediate rule. Similarly, the scale rates s_j and the move ratios $\mathbb{M}_j$ $(j = 1, \ldots, M)$ are combined to s' and $\mathbb{M}'$ by using the function f_7.

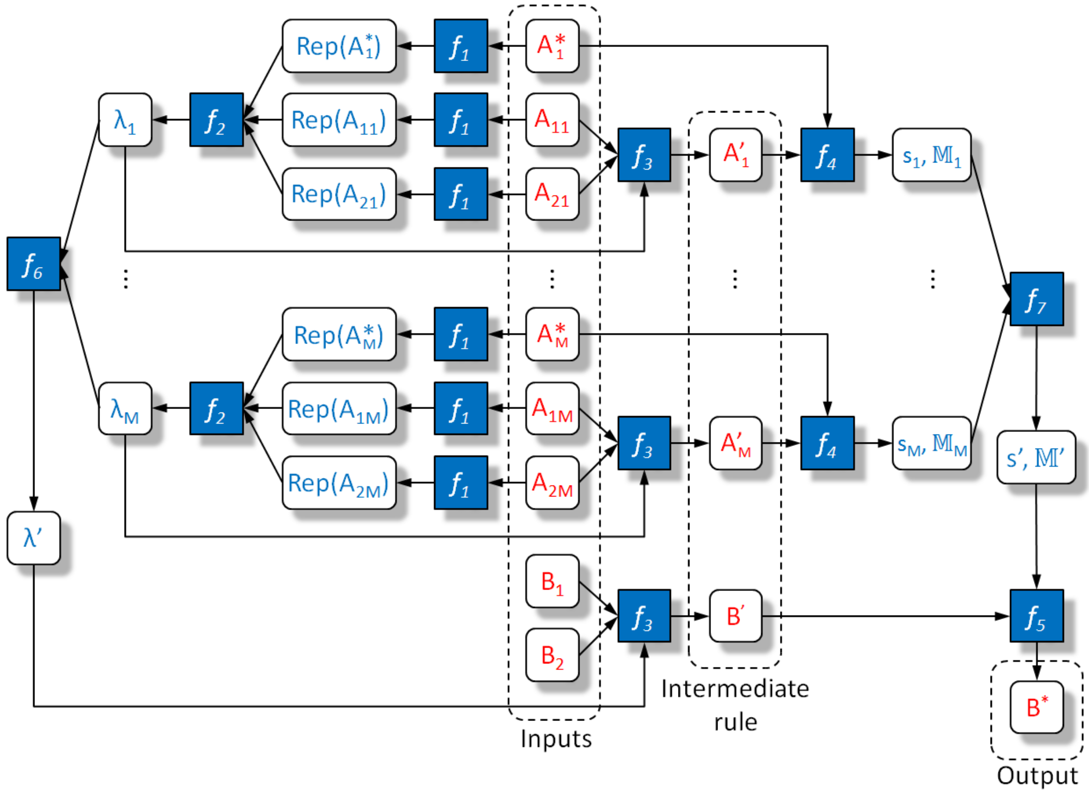


Figure 2.7: T-FRI with two multi-antecedent rules

2.3.2.4 The T-FRI Approach with Multiple Multi-antecedent Rules

In order to implement interpolation or extrapolation with multiple multi-antecedent rules, the first step is to choose N $(N \geq 2)$ rules from a given rule base. Then, an intermediate rule is constructed based on the selected rules. Once the intermediate rule is worked out, the remainder of the process remains the same as that described in the previous sections. The key steps in generating an intermediate rule are briefly introduced as follows.

Closest N Rules Selection

Without loss of generality, suppose that a rule R_i and an observation O are represented by

$$R_i : \text{If } x_1 \text{ is } A_{i1}, \cdots, x_j \text{ is } A_{ij}, \cdots, x_M \text{ is } A_{iM}, \text{ then } y \text{ is } B_i$$
$$O : x_1 \text{ is } A_1^*, \cdots, x_j \text{ is } A_j^*, \cdots, x_M \text{ is } A_M^*$$

where A_{ij} denotes the jth antecedent FS of Rule R_i, A^*_j denotes the observed FS of variable x_j, and B_i denotes the consequent FS of Rule R_i with $j \in \{1,\ldots,M\}$, M being the number of antecedent variables.

The distances d_{ij} between the pairs of A_{ij} and A^*_j can be calculated as follows:

$$\begin{aligned} d_{ij} &= d(A_{ij}, A^*_j) \\ &= d(\text{Rep}(A_{ij}), \text{Rep}(A^*_j)) \end{aligned} \tag{2.44}$$

The distance d_i between the rule R_i and the observation O is deemed to be the average of all antecedent variables' distances:

$$d'_{ij} = \frac{d_{ij}}{\max_j - \min_j} \tag{2.45}$$

$$d_i = \sqrt{\sum_{j=1}^{M} {d'_{ij}}^2} \tag{2.46}$$

where $\max_j$ and $\min_j$ are the maximum and minimum values of x_j, $j \in \{1,\ldots,M\}$. Each distance measure d_{ij} is normalised into the range $[0,1]$, denoted by d'_{ij}, to make the absolute distances compatible with each other over different domains. Note that if $\max_j - \min_j = 0$, then $\max_j = \min_j$. That is, A^*_j of O is identical with A_{ij} of R_i, $j \in \{1,\ldots,M\}$. In this case, $d'_{ij} = 1$.

Intermediate Rule Construction

Suppose N $(N \geq 2)$ closest rules have been chosen from the observation. Such rules are represented as R_i, $i \in \{1,\ldots,N\}$, each has M antecedents A_{ij}, $j \in \{1,\ldots,M\}$. Let $w_{A_{ij}}$ denote the weight to which the jth antecedent of the ith rule contributes to the intermediate rule. The normalised weight $w'_{A_{ij}}$ can be defined as:

$$w_{A_{ij}} = \frac{1}{d_{ij}} \tag{2.47}$$

$$w'_{A_{ij}} = \frac{w_{A_{ij}}}{\sum_{i=1}^{N} w_{A_{ij}}} \tag{2.48}$$

Note that if $d_{ij} = 0$, then $\text{Rep}(A_{ij}) = \text{Rep}(A^*_j)$. In this case, the antecedent of the observation is considered to be identical to the corresponding antecedent of the rule

R_i, in terms of the currently applied definition of Rep. Thus, $w_{A_{ij}} = 1$ for the identical ones, while $w_{A_{ij}} = 0$ for the remainder.

The antecedent of the so-called intermediate fuzzy term A_j^{IFT} is constructed from the antecedents of these closest rules. Another process *shift* is then introduced to modify A_j^{IFT} to the antecedent of the intermediate rule A'_j so that it will have the same Rep as A_j^*:

$$A_j^{IFT} = \sum_{i=1}^{N} w'_{A_{ij}} A_{ij} \tag{2.49}$$

$$A'_j = A_j^{IFT} + \delta_{A_j}(\max_j - \min_j) \tag{2.50}$$

where δ_{A_j} is a constant defined by

$$\delta_{A_j} = \frac{\mathrm{Rep}(A_j^*) - \mathrm{Rep}(A_j^{IFT})}{\max_j - \min_j} \tag{2.51}$$

Note that if $\max_j - \min_j = 0$, then $\max_j = \min_j$. That is, A_j^* is identical with A_j^{IFT}, $j \in \{1, \ldots, M\}$. In this case, $\delta_{A_j} = 1$. Regarding the consequence of the intermediate rule B', it can be calculated by analogy to the computation of the antecedent, such that

$$B^{IFT} = \sum_{i=1}^{N} w'_{B_i} B_i \tag{2.52}$$

$$B' = B^{IFT} + \delta_B(\max - \min) \tag{2.53}$$

where B^{IFT} is the consequence of the intermediate fuzzy term, max and min are the maximum and minimum values of consequent variable, w'_{B_i} and δ_B are the means of $w'_{A_{ij}}$ and δ_{A_j}, $i \in \{1, \ldots, N\}$, $j \in \{1, \ldots, M\}$, respectively, which are defined as:

$$w'_{B_i} = \frac{1}{M} \sum_{j=1}^{M} w'_{A_{ij}} \tag{2.54}$$

$$\delta_B = \frac{1}{M} \sum_{j=1}^{M} \delta_{A_j} \tag{2.55}$$

Then, the intermediate rule is constructed as

$$\text{If } x_1 \text{ is } A'_1, \cdots, x_j \text{ is } A'_j, \cdots, x_M \text{ is } A'_M, \text{ then } y \text{ is } B'.$$

Having generated the required intermediate rule, the rest of the interpolation involves firing this rule by the given observation, which is the same as that of interpolation with two rules described previously. The process of the T-FRI with multiple multi-antecedent rules is illustrated in Figure 2.8. In addition, extrapolation is a special case of interpolation when all the N closest rules lie on one side of the given observation. However, the processes of choosing the closest rules and constructing the intermediate rule are carried out in exactly the same way as the procedures for interpolation.

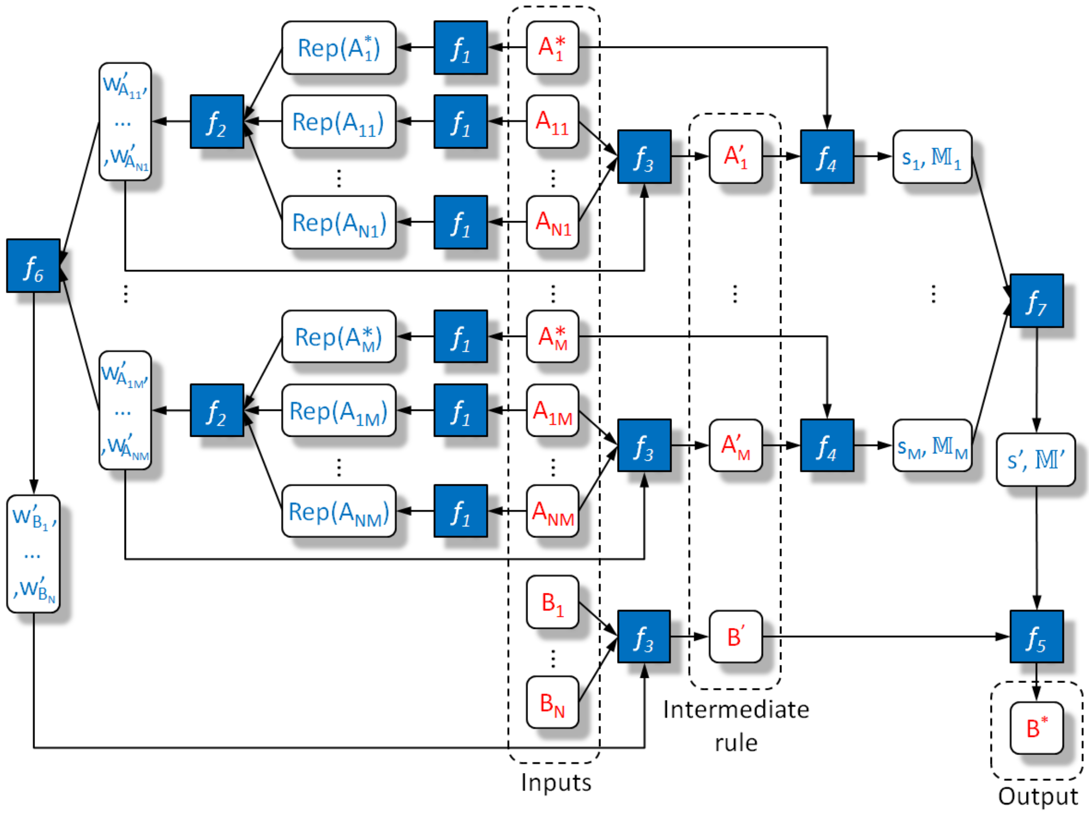


Figure 2.8: T-FRI with multiple multi-antecedent rules

2.3.3 Other Approaches

In addition to the aforementioned approaches, a number of other existing approaches have also been reported in the literature [23, 39, 73, 84, 85, 91, 111, 157], several of them are reviewed in the following sections. For details of other implementations, refer to the corresponding references given.

2.3.3.1 HCL Interpolation

The HCL (Hsiao, Chen, and Lee) approach [73] eliminates the abnormal problem by fixing the core of the consequence generated by the KH approach and shifting its support along with the consequent variable axis. It represents both *slopes* of each FS as a linear function. The slopes of the consequent FS are also linear functions whose parameters are interpolated from those of the observation and the FSs involved in the rule bases. A ratio between the *left slope* and the *right slope* of the consequence is then calculated and utilised to shift the support of the generated consequence by the KH approach in reference to the normal point of the consequence. Unfortunately, this approach is only applicable to triangular FSs.

The typical interpolation problem is shown in Figure 2.9, where k_1, t_1, k, t, k_2, t_2, h_1, m_1, h, m, h_2, and m_2 represent the slopes of the corresponding FSs. The HCL approach calculates the support of B^* in the same way as the KH approach but the top point is calculated in a different way. The process to determine the top point of B^* is described below.

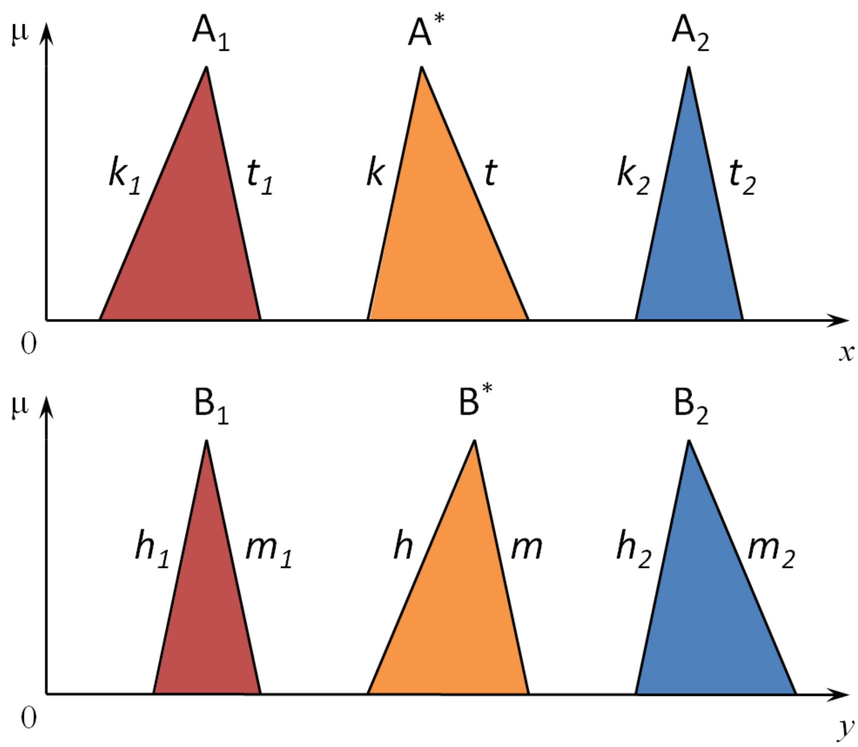


Figure 2.9: HCL interpolation

The slopes h and m of B^* are calculated first. Let:

$$\begin{cases} k = k_1 x + k_2 y \\ t = t_1 x + t_2 y \end{cases} \tag{2.56}$$

where x and y are real numbers. If $\frac{k_1}{t_1} \neq \frac{k_2}{t_2}$, then x and y are computed by Equation (2.56). Let:

$$\begin{cases} h = c|h_1 x + h_2 y| \\ m = -c|m_1 x + m_2 y| \end{cases} \tag{2.57}$$

where c is a constant. Otherwise, let:

$$\begin{cases} h = ck \\ m = ct \end{cases} \tag{2.58}$$

where c is a constant.

The position of the top point of B^* is then decided by

$$\frac{\mathrm{CP}(B^*) - \sup(B^*)}{\mathrm{CP}(B^*) - \inf(B^*)} = \frac{h}{m} \tag{2.59}$$

where $\mathrm{CP}(B^*)$ denotes the centre point of core of B^*. Equation (2.59) can be reformulated as

$$\mathrm{CP}(B^*) = \frac{\sup(B^*)m - \inf(B^*)h}{m - h} \tag{2.60}$$

Note that if $m = h$, then $\sup(B^*) = \inf(B^*)$ can be derived from Equation (2.59). In this case, $\mathrm{CP}(B^*) = \sup(B^*) = \inf(B^*)$.

2.3.3.2 CCL Interpolation

The CCL (Chang, Chen, and Liau) approach [23] can be seen as an improvement of the HCL approach. This approach first determines the core of the consequence by using the KH approach, which is calculated as follows:

$$b^* = b_1 + \frac{(a^* - a_1)(b_2 - b_1)}{a_2 - a_1} \tag{2.61}$$

where a_1, a_2, a^*, b_1, b_2, and b^* are the normal points of the involved triangular FSs A_1, A_2, A^*, B_1, B_2, and B^*, respectively.

The areas of the two sides of the core are then calculated from the corresponding areas of the given observation and all the FSs involved in the rules used for interpolation in a manner of linear interpolation.

$$S_K(B^*) = \begin{cases} S_K(A^*) \sum_{i=1}^{2} W_i \dfrac{S_K(B_i)}{S_K(A_i)} & \text{if } \exists i\ S_K(A_i) > 0 \\ S_K(A^*) & \text{if } \forall i\ S_K(A_i) = 0 \end{cases} \tag{2.62}$$

where $K \in \{L, R\}$, $S_L(B^*)$ and $S_R(B^*)$ denote the left and the right area of B^*, respectively, and

$$\begin{cases} W_1 = 1 - \dfrac{a^* - a_1}{a_2 - a_1} \\ W_2 = 1 - W_1 \end{cases} \tag{2.63}$$

The interpolated result B^* is therefore derived by

$$B^* = (b^* - 2S_L(B^*), b^*, b^* + 2S_R(B^*)) \tag{2.64}$$

Unlike the HCL approach, this approach is able to deal with interpolation and extrapolation with multiple multi-antecedent rules, with each rule involving any shape of FSs.

2.3.3.3 QMY Interpolation

The QMY (Qiao, Mizumoto, and Yan) approach [157] employs the same mechanism for generating intermediate rules as the T-FRI approach, but the Rep is restricted to being the centre point of *core*. The similarity degree between the observation A^* and the antecedent A' of the intermediate rule is captured using the so-called parameters *lower similarity* and *upper similarity*, which are defined by

$$\begin{cases} S_{L(A^*,A')}(\alpha) = \dfrac{d(\inf(A^*_\alpha), \mathrm{CP}(A^*))}{d(\inf(A'_\alpha), \mathrm{CP}(A^*))} \\ S_{U(A^*,A')}(\alpha) = \dfrac{d(\sup(A^*_\alpha), \mathrm{CP}(A^*))}{d(\sup(A'_\alpha), \mathrm{CP}(A^*))} \end{cases} \tag{2.65}$$

where $\alpha \in (0, 1]$.

With reference to the centre point of the core, a convex and normal FS can be divided into two parts, namely, the lower part and the upper part. The lower similarity measures the difference of the lower parts of two FSs, by comparing the lengths of a certain level cut, and the upper similarity does that of the upper parts.

In so doing, the consequence B^* is derived from the following equations:

$$\begin{cases} \mathrm{CP}(B^*) = \mathrm{CP}(B') \\ S_{L(B^*,B')}(\alpha) = S_{L(A^*,A')}(\alpha) \\ S_{U(B^*,B')}(\alpha) = S_{U(A^*,A')}(\alpha) \end{cases} \tag{2.66}$$

Combining Equations (2.65) and (2.66) gives

$$\begin{cases} \inf(B^*_\alpha) = S_{L(A^*,A')}(\alpha)d(\inf(B'_\alpha), \mathrm{CP}(B')) + \mathrm{CP}(B') \\ \sup(B^*_\alpha) = S_{U(A^*,A')}(\alpha)d(\sup(B'_\alpha), \mathrm{CP}(B')) + \mathrm{CP}(B') \end{cases} \tag{2.67}$$

Thus B^* can be calculated with the representation principle of FSs.

2.3.3.4 CK Interpolation

The CK (Chen and Ko) approach [39] ensures that the *core* of each FS of a created intermediate rule is equal to that of the corresponding FS of the resultant interpolated rule.

First, the Reps of all the involved FSs are obtained by the T-FRI approach, resulting in the parameter λ. The values of $la'_{0,1}$ and $la'_{1,2}$ are then calculated:

$$\begin{cases} la'_{0,1} = & (1-\lambda)la_{1_{0,1}} + \lambda la_{2_{0,1}} \\ la'_{1,2} = & (1-\lambda)la_{1_{1,2}} + \lambda la_{2_{1,2}} \end{cases} \tag{2.68}$$

where $la'_{0,1}$ and $la'_{1,2}$ denote the left and the right support length of the antecedent of the intermediate rule. The values of $lb'_{0,1}$ and $lb'_{1,2}$ can be calculated in the same way.

Next, the antecedent of the intermediate rule is constructed:

$$\begin{cases} a'_0 = a'_1 - la'_{0,1} \\ a'_1 = a_1 \\ a'_2 = a'_1 + la'_{1,2} \end{cases} \tag{2.69}$$

Similarly, the consequence of the intermediate rule can be constructed by means of the previously obtained $lb'_{0,1}$ and $lb'_{1,2}$

$$\begin{cases} b'_0 = b'_1 - lb'_{0,1} \\ b'_1 = b_1 \\ b'_2 = b'_1 + lb'_{1,2} \end{cases} \tag{2.70}$$

where b_1 is the core of the estimated interpolated conclusion, which is determined as follows:

$$b_1 = (1 - \lambda_{a_1})\mathrm{Rep}(B_1) + \lambda_{a_1}\mathrm{Rep}(B_2) \tag{2.71}$$

where

$$\begin{aligned} \lambda_{a_1} &= \frac{d(A_1, a_1)}{d(A_1, A_2)} \\ &= \frac{a_1 - \mathrm{Rep}(A_1)}{\mathrm{Rep}(A_2) - \mathrm{Rep}(A_1)} \end{aligned} \tag{2.72}$$

Note that $\mathrm{Rep}(A_1) \neq \mathrm{Rep}(A_2)$ because $A_1 \prec A_2$ or $A_2 \prec A_1$.

In order to measure the similarity degree between two FSs with the same core, only their *left slopes* and *right slopes* need to be compared. Two transformations, i.e., *increment transformation* and *ratio transformation* are then utilised for this purpose, with one aiming to increase the length of a certain level cut of a *slope* during the transformation, and the other to decrease the length. From this, $B^* = (b^*_0, b^*_1, b^*_2)$ can be derived, where b_0 and b_2 are calculated as

$$b_0 = \begin{cases} b_1 - la_{0,1}\dfrac{a_{20} - a_{12}}{b_{20} - b_{12}} + la'_{0,1}\dfrac{a_{20} - a_{12}}{b_{20} - b_{12}} - lb'_{0,1} & \text{if } la_{0,1} \geq la'_{0,1} \\ b_1 - \dfrac{la_{0,1} lb'_{0,1}}{la'_{0,1}} & \text{otherwise} \end{cases} \tag{2.73}$$

$$b_2 = \begin{cases} b_1 + la_{1,2}\dfrac{a_{20} - a_{12}}{b_{20} - b_{12}} - la'_{1,2}\dfrac{a_{20} - a_{12}}{b_{20} - b_{12}} + lb'_{1,2} & \text{if } la_{1,2} \geq la'_{1,2} \\ b_1 + \dfrac{la_{1,2} lb'_{1,2}}{la'_{1,2}} & \text{otherwise} \end{cases} \tag{2.74}$$

2.4 Knowledge Representation

Fuzziness differs from generality, vagueness, and ambiguity in that it is not simply a result of a one-to-many relationship between a general meaning and its specifications; nor a list of possible related interpretations derived from a vague expression

[115]; nor a list of unrelated meanings denoted by an ambiguous expression [224]. Fuzziness is inherent in the sense that it measures the degree to which an event occurs. It is explored for describing uncertainty.

Humans and machines represent their knowledge in many different ways and formats, and this knowledge is often vague, ambiguous and incomplete. Efficient communication of knowledge relies on an understanding of the representation of uncertain information and knowledge in the problem domain [16, 22, 166]. When knowledge is represented as a set of facts and rules, this uncertainty can be measured by means of a number of different approaches, including those to be outlined below as well as given previously.

2.4.1 Rough Set Theory

Dealing with incomplete or imperfect knowledge lies outside the core of much research in computational intelligence and cognitive sciences. Being able to understand and manipulate such knowledge is of fundamental significance to many theoretical developments and practical applications of automation and computing [88], particularly in the areas of decision analysis, machine learning and data-mining, intelligent control, and pattern recognition. RS theory [50, 51, 151, 152, 153] offers one of the most distinct and recent approaches for modelling imperfect knowledge. Owing to the recognition of the existing and potentially important impact of this theory, it has attracted worldwide attention of further research and development, resulting in various extensions to the original theory and increasingly widening fields of application [66, 86, 87, 89, 148, 149].

2.4.1.1 Information Systems

A data set can be represented as a table, where each row represents an *object* (a case, an event, a person, etc.). Each column represents an *attribute* (a variable, an observation, a property, etc.) that can be measured for each data object. The attribute may be also supplied by a human expert or user. This table is called an *information system (information table)* [110], as shown in Table 2.1.

An information system may be extended by the inclusion of decision attributes. Such a system is called a *decision system (decision table)*, columns of which are labelled *attributes*, rows - by *objects* of interest and entries of the table are *attribute*

Table 2.1: An example information system

	Attributes			**Decision**
	Headache	Muscle Pain	Temperature	Flu
p1	yes	yes	normal	no
p2	yes	yes	high	yes
p3	yes	yes	very high	yes
p4	no	yes	normal	no
p5	no	no	high	no
p6	no	yes	very high	yes

values. Attributes of the decision system are divided into two disjoint groups called *condition* and *decision* attributes, respectively. A decision system is *consistent* if for every set of objects whose attribute values are the same, the corresponding decision attributes are identical [150, 154].

More formally, $I = (\mathbb{U}, \mathbb{A})$ is an information system, where $\mathbb{U}$ is a non-empty set of finite objects (the universe of discourse) and $\mathbb{A}$ is a non-empty finite set of attributes such that $a : \mathbb{U} \rightarrow V_a$ for every $a \in \mathbb{A}$. V_a is the set of values that attribute a may take. For decision systems, $\mathbb{A} = \mathbb{C} \cup \mathbb{D}$, where $\mathbb{C}$ is the set of input features and $\mathbb{D}$ is the set of class indexes. Here, a class index $d \in \mathbb{D}$ is itself a variable $d : \mathbb{U} \rightarrow \{0, 1\}$ such that for $a \in \mathbb{U}$, $d(a) = 1$ if a has class d and $d(a) = 0$ otherwise [88].

2.4.1.2 Indiscernibility

RS theory is founded on the assumption that with every object of the universe of discourse, some information (data, knowledge) is associated with it. Objects characterised by the same information are *indiscernible* in view of the available information about them. The *indiscernibility relation* generated in this way forms the mathematical basis of RS theory.

Let $I = (\mathbb{U}, \mathbb{A})$ be an information system, then with any $P \subseteq \mathbb{A}$ there is a crisp equivalence relation $IND(P)$:

$$IND(P) = \{(x, y) \in \mathbb{U}^2 \mid \forall a \in P,\ a(x) = a(y)\} \tag{2.75}$$

If $(x, y) \in IND(P)$, then x and y are indiscernible by attributes from P. The equivalence class with respect to such an indiscernibility relation defined on P is denoted by $[x]_P$, $x \in \mathbb{U}$.

Any set of all *indiscernible* objects is called an *elementary set (concept)*, and forms a basic *granule (atom)* of knowledge about the universe. Any union of some elementary sets is referred to as a *crisp (precise)* set - otherwise the set is *rough (imprecise, vague)*.

2.4.1.3 Lower and Upper Approximations

Let $X \subseteq \mathbb{U}$, X can be approximated using only the information contained within P by constructing the P-*lower* and P-*upper* approximations of X:

$$\begin{cases} \underline{P}X = \{x \mid [x]_P \subseteq X\} \\ \overline{P}X = \{x \mid [x]_P \cap X \neq \emptyset\} \end{cases} \tag{2.76}$$

The tuple $< \underline{P}X, \overline{P}X >$ is called an RS.

Consider the approximation of concept X in Figure 2.10. Each square or granule in the diagram represents an equivalence class, generated by indiscernibility between object values. Using the features in set P, via these equivalence classes, the *lower* and *upper* approximations of X can be constructed.

Equivalence classes contained within X belong to the lower approximation (LA). Objects lying within this region can be said to *certainly* belong to concept X. Equivalence classes within X and along its border form the upper approximation (UA). Those objects in this region can only be said to *possibly* belong to the concept. The difference between the LA and the UA constitutes the *boundary region* of the RS.

2.4.1.4 Positive, Negative and Boundary Regions

Let P and Q be sets of attributes inducing equivalence relations over $\mathbb{U}$, then the positive, negative, and boundary regions are defined as follows:

$$\begin{aligned} POS_P(Q) &= \bigcup_{X \in \mathbb{U}/Q} \underline{P}X \\ NEG_P(Q) &= \mathbb{U} - \bigcup_{X \in \mathbb{U}/Q} \overline{P}X \\ BND_P(Q) &= \bigcup_{X \in \mathbb{U}/Q} \overline{P}X - \bigcup_{X \in \mathbb{U}/Q} \underline{P}X \end{aligned} \tag{2.77}$$

where $\mathbb{U}/Q$ is defined as the equivalence classes of the relation $IND(Q)$.

The positive region, $POS_P(Q)$, comprises all objects of $\mathbb{U}$ that can be classified to classes of $\mathbb{U}/Q$ using the information contained within attributes P. The negative

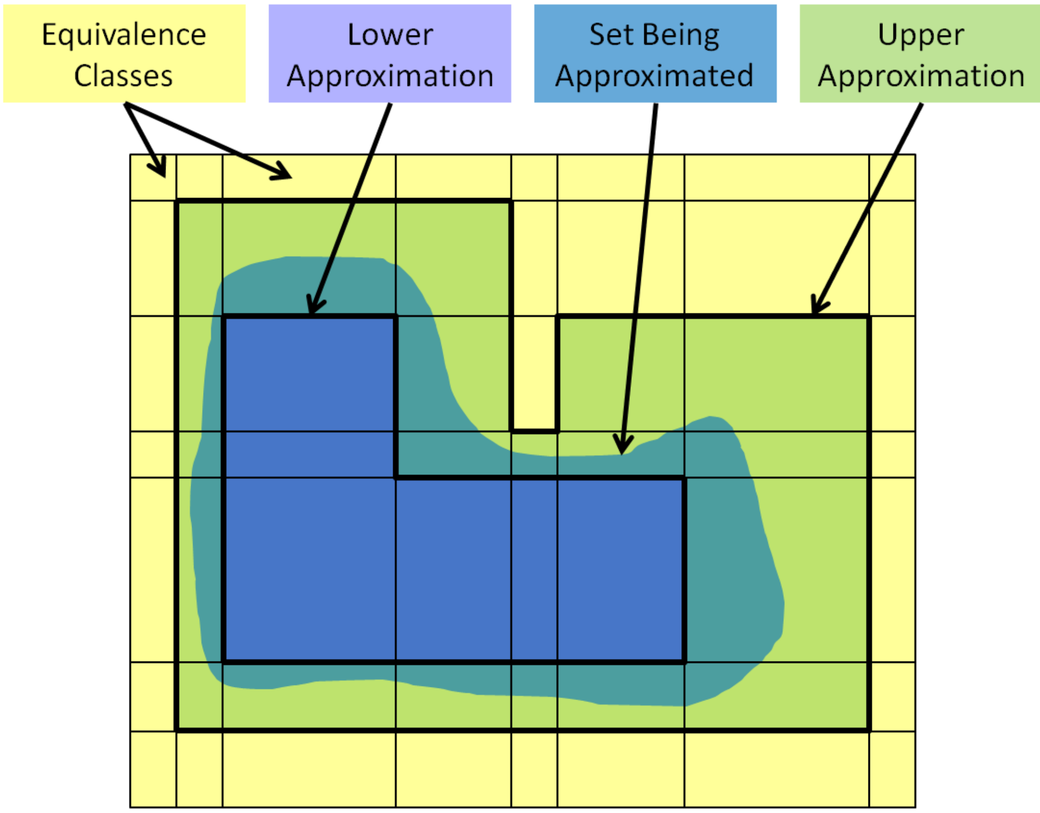


Figure 2.10: Basic concepts of RS

region, $NEG_P(Q)$, is the set of objects that cannot be classified to classes of $\mathbb{U}/Q$. The boundary region, $BND_P(Q)$, is the set of objects that can possibly, but not certainly, be classified in this way.

If the boundary region is the empty set, i.e., $BND_P(Q) = \emptyset$, then X is *crisp* with respect to P. In the opposite case, i.e., if $BND_P(Q) \neq \emptyset$, X is referred to as *rough* with respect to P.

2.4.2 Type-2 Fuzzy Set Theory

Type-2 FSs were first defined and discussed in [215], this work concentrated on the notion of an FS where the membership grades of an FS are measured with linguistic terms such as *low*, *medium* and *high* [94, 213]. In other words, a conventional (type-1) FS has a grade of membership that is crisp, whereas a type-2 FS has grades of membership that are fuzzy, so it could be called a "fuzzy-fuzzy set" [135]. Hence, the MF of a type-2 FS is three-dimensional, and it is the third dimension that provides a new degree of freedom for handling uncertainty [133]. Such sets are useful in situations where there is uncertainty about the membership grades themselves, e.g., an uncertainty in the shape of the MF or in some of its parameters. Consider the transition from ordinary sets to FSs. When the membership of an element in a set cannot be determined by 0 or 1, type-1 FSs are used. Similarly, when the circumstances are so fuzzy that determining the membership grade is difficult even as a crisp number in $[0,1]$, type-2 FSs are then required [100].

A view of the relationships between levels of imprecision, data and techniques is shown in Figure 2.11. As the level of imprecision increases, type-2 fuzzy logic provides a powerful paradigm for handling the problem. Problems that contain crisp, precise data do not, in reality, exist. However, some problems can be solved effectively with mathematical techniques where the assumption is that the data is precise. Other problems use imprecise terminology that can often be effectively modelled by using type-1 FSs. Here, perceptions are at a higher level of imprecision and type-2 FSs can effectively model this imprecision [94].

2.4.2.1 Definitions

A type-2 FS is characterised by a fuzzy MF whose membership grade for each element is a fuzzy number in $[0,1]$. The formed definition is provided below.

Definition 2.1. [137] A *type-2 FS*, denoted $\tilde{A}$, is characterised by a *type-2 MF* $\mu_{\tilde{A}}(x,u)$, where $x \in X$ and $u \in J_x \subseteq [0,1]$, i.e.,

$$\tilde{A} = \{((x,u), \mu_{\tilde{A}}(x,u)) | \forall x \in X, \forall u \in J_x \subseteq [0,1]\} \tag{2.78}$$

where J_x is the primary membership of x, $J_x = \{u : \mu_{\tilde{A}}(x,u) \subseteq [0,1]\}$. $\tilde{A}$ can also be expressed as:

$$\tilde{A} = \int_{x \in X} \int_{u \in J_x} \frac{\mu_{\tilde{A}}(x,u)}{(x,u)}, \quad J_x \subseteq [0,1] \tag{2.79}$$

where $\int \int$ denotes union over all admissible x and u.

In general, a *type-2 FS* is referred to as a *general type-2 FS* in order to distinguish it from the special interval type-2 FS, which is defined below.

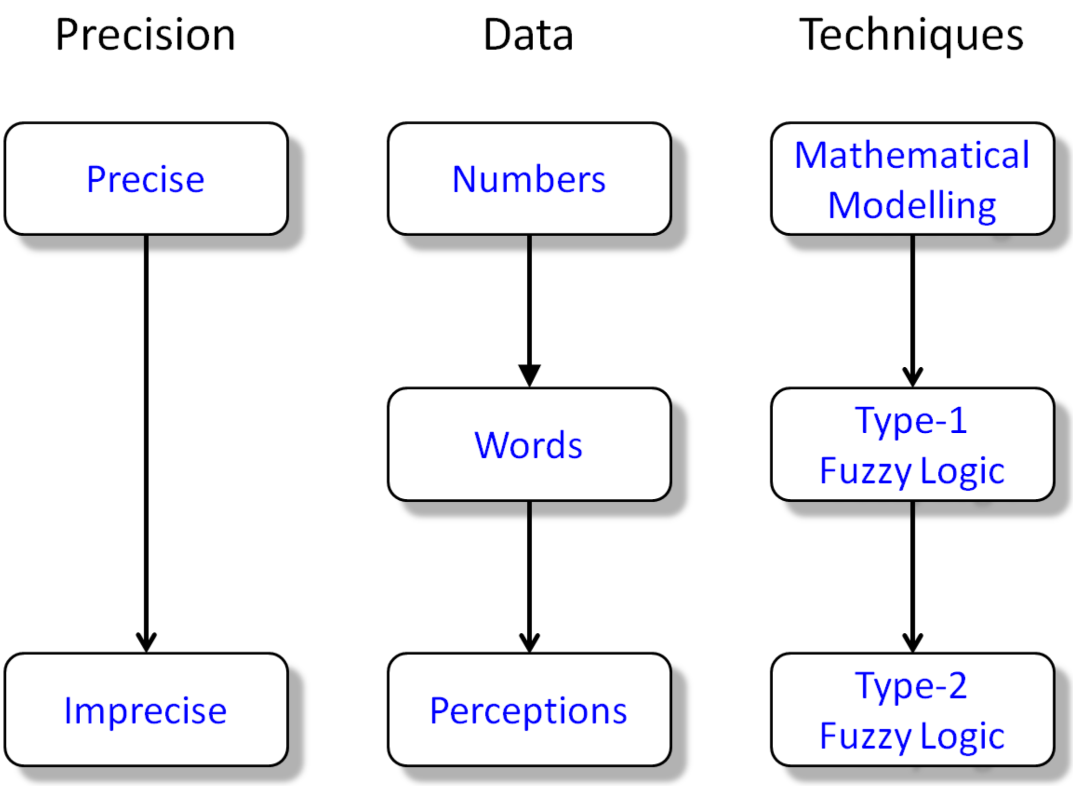


Figure 2.11: Relationships between imprecision, data and techniques

Definition 2.2. [136] When all $\mu_{\tilde{A}}(x,u) = 1$ then $\tilde{A}$ is an *interval type-2 FS*, which is expressed as:

$$\tilde{A} = \int_{x \in X} \int_{u \in J_x} \frac{1}{(x,u)}, \quad J_x \subseteq [0,1] \tag{2.80}$$

Definition 2.3. [136] Uncertainty in the primary memberships of an interval type-2 FS, $\tilde{A}$, consists of a bounded region that is called the *footprint of uncertainty* (FOU). It is the union of all primary memberships, i.e.,

$$\mathrm{FOU}(\tilde{A}) = \bigcup_{x \in X} J_x \tag{2.81}$$

Definition 2.4. [136] The *upper MF* and *lower MF* of $\tilde{A}$ are two type-1 MFs that bound the FOU. The upper MF is associated with the upper bound of FOU($\tilde{A}$) and is denoted $\overline{\mu}_{\tilde{A}}(x)$, $\forall x \in X$, and the lower MF is associated with the lower bound of

FOU($\tilde{A}$) and is denoted $\underline{\mu}_{\tilde{A}}(x)$, $\forall x \in X$, i.e.,

$$\begin{cases} \overline{\mu}_{\tilde{A}}(x) = \overline{\mathrm{FOU}(\tilde{A})} = \sup J_x, & \forall x \in X \\ \underline{\mu}_{\tilde{A}}(x) = \underline{\mathrm{FOU}(\tilde{A})} = \inf J_x, & \forall x \in X \end{cases} \tag{2.82}$$

For an interval type-2 FS, its third-dimension value is the same everywhere which means that no new information is contained in the third dimension. In this case, the third dimension is then ignored, and only the FOU is used to describe such a set, which is shown in Figure 2.12. Such an interval type-2 FS is completely characterised by its FOU that is bounded by lower MF and upper MF, and, its embedded FSs are type-1 FSs.

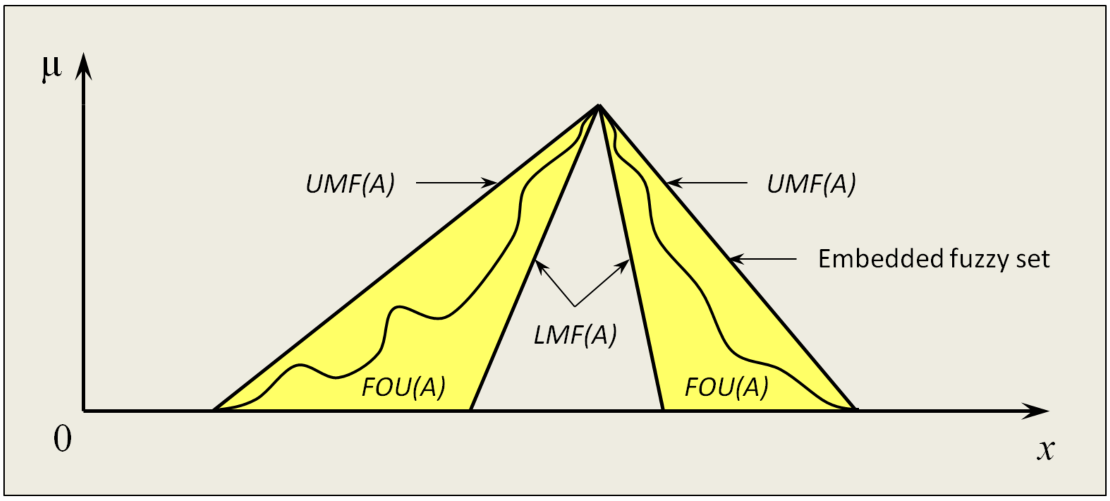


Figure 2.12: An interval type-2 FS

2.4.2.2 Type-2 Fuzzy Logic Systems

A fuzzy logic system (FLS) [101, 122] (also known as FIS, fuzzy controller, etc.) includes fuzzifier, rule base, inference engine, and defuzzifier. Quite often, the knowledge used to construct rules in an FLS is uncertain. This uncertainty leads to rules having uncertain antecedents and/or consequences, which in turn translates into uncertain antecedent and/or consequent MFs.

Basically, there are (at least) four types of uncertainty in type-1 FLSs [137]: (1) The meanings of the words that are used in the antecedents and consequences of rules can be uncertain (words mean different things to different people). (2) Consequences may have a histogram of values associated with them, especially when knowledge is extracted from a group of experts who do not all agree. (3) Measurements that activate a type-1 FLS may be noisy and therefore uncertain. (4) The data that are used to tune the parameters of a type-1 FLS may also be noisy.

Most of these types of uncertainty translate into difficulties about FS MFs. Type-1 FSs are not able to model such types of uncertainty because their MFs are crisp. On the contrary, type-2 FSs are able to model such uncertainty, because their MFs are themselves fuzzy.

The structure of a type-2 FLS is very similar to the structure of a type-1 FLS, which is shown in Figure 2.13. A type-2 FLS is characterised by IF-THEN rules, but its antecedent and/or consequent sets are now type-2 FSs. It includes fuzzifier, rule base, inference engine, and output processing. For a type-1 FLS, the *output processing* block only contains the defuzzifier.

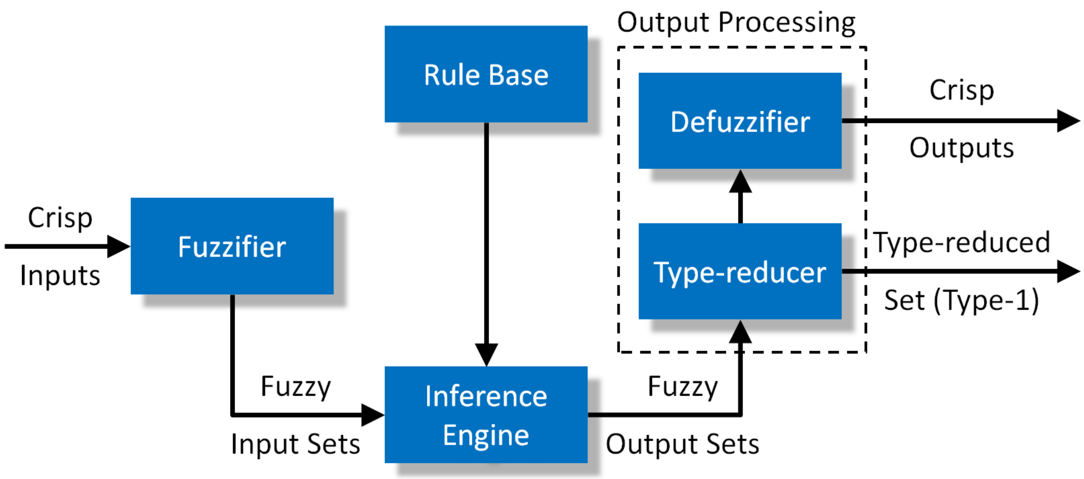


Figure 2.13: Type-2 FLSs

The *fuzzifier* maps the crisp input into an FS. In general, this FS can be a type-2 set or a singleton where the input FS only has a single point of non-zero membership.

For the *rule base*, the distinction between type-1 and type-2 is associated with the nature of the MFs, which is not important while forming rules. For this reason, the structure of the rules remains exactly the same in type-2 FLSs, the only difference being that some or all of the involved sets are of type-2. However, it is not necessary that all the antecedents and consequences be type-2 FSs. As long as one antecedent or the consequent set is type-2, it is a type-2 FLS.

The inference engine in a type-1 FLS combines rules and gives a mapping from input type-1 FSs to output type-1 FSs. Multiple antecedents in rules and multiple rules are connected by the *t*-norm (corresponding to intersection of sets) and the *t*-conorm (corresponding to the union of sets), respectively. Similarly, the *inference engine* in a type-2 FLS combines rules and gives a mapping from input type-2 FSs to output type-2 FSs with the use of intersections and unions of type-2 FSs.

In a type-1 FLS, the defuzzifier produces a crisp output from the FS that is the output of the inference engine, i.e., a type-0 (crisp) output is obtained from a type-1 set. In the type-2 case, an operation analogous to type-1 defuzzification results in a type-1 set from a type-2 set, which is the output of the inference engine. This operator is called *type-reducer* and the resultant set is called a “type-reduced set”. This type-reduced set can be further defuzzified by the *defuzzifier* to obtain a crisp output. The most natural way of doing this seems to be by finding the centroid of the type-reduced set [99, 125], however, there exist other possibilities like choosing the highest membership point in the type-reduced set [44, 184, 185].

2.5 Aggregation Techniques

The aggregation and fusion of information are basic concerns [47] for all kinds of knowledge-based systems [1, 45], from image processing [82, 160] to decision making [8, 223], from pattern recognition [9, 11] to machine learning [2, 141]. Information aggregation is a process in which information is gathered and expressed in a summarised form. The goal of aggregation is to integrate and refine information resulting from various sources, in order to form a better conclusion or decision than from individual sources only, by reducing imprecision and uncertainty while increasing completeness.

Informally, an aggregation process involves combining an n-tuple of objects all concerning a given concept into a single object regarding the same concept. In the case of mathematical aggregation, an aggregation operator is typically a function, which assigns a real number y to any n-tuple $(x_1, x_2, \cdots, x_n)$ of real numbers [47]:

$$y = Agg(x_1, x_2, \cdots, x_n) \tag{2.83}$$

More generally, aggregation operators are mathematical functions that reduce a set of numbers into a unique representative number. There are a number of well-known aggregation operators. For instance, the simplest and most common way to aggregate is to use the *arithmetic mean*, and also the *weighted mean*, which allows placing weights on the arguments to be averaged. In addition, the *minimum*, the *maximum* and the *median* are also commonly used aggregation operators [190].

2.5.1 OWA-based Aggregation

Apart from the aforementioned classical aggregation operators, a new information aggregation technique was proposed based on the OWA scheme [191, 194, 195]. The OWA operator considers a wide range of averaging operators that move between the minimum and the maximum. It allows the aggregation of information in considering the degree of optimism or pessimism that a decision maker wants to express in the aggregation itself. OWA-based aggregation strategies have been widely investigated and have achieved successful applications in many different domains, such as decision making [70, 193], fuzzy control [198, 200], market analysis [201], image compression [140], etc.

Definition 2.5. [191] An OWA operator of dimension n is a mapping $\mathbb{R}^n \rightarrow \mathbb{R}$, which has an associated weighting vector $W = (w_1, w_2, \ldots, w_n)^T$, where $w_i \in [0, 1]$ and $\sum_{i=1}^{n} w_i = 1$. An input vector $(a_1, a_2, \ldots, a_n)$, is aggregated as follows:

$$\mathrm{OWA}(a_1, a_2, \ldots, a_n) = \sum_{i=1}^{n} w_i b_i \tag{2.84}$$

where b_i is the ith largest element in the vector $(a_1, a_2, \ldots, a_n)$ and $b_1 \geq b_2 \geq \cdots \geq b_n$.

Generally speaking, the OWA-based aggregation process consists of three steps:

- Reorder the input arguments in descending order.
- Determine the weights associated with the OWA operator by using a proper method.
- Utilise the OWA weights to aggregate these reordered arguments.

A fundamental aspect of the OWA operator is the reordering step, which makes this a non-linear operator. During this step the arguments are ordered by their values. In particular, the weights rather than being associated with a specific argument, as in the case of the usual weighted average, are associated with a particular position in the ordering. Clearly, one key point required to implement the OWA operators is to determine the associated weights. In general, different choices of the weight vector W lead to different aggregation results [59, 186, 192]. Actually, the OWA operators provide a parameterised family of mean type aggregation operators, which include many of the classical operators. Several particular cases of the OWA operators are listed in Table 2.2.

Table 2.2: Particular cases of OWA

	OWA
Arithmetic mean	$w_i = \frac{1}{n}$ for $\forall i$
Minimum	$\begin{cases} w_1 = 1 \\ w_i = 0 \text{ if } i \neq 1 \end{cases}$
Maximum	$\begin{cases} w_n = 1 \\ w_i = 0 \text{ if } i \neq n \end{cases}$
Median	$\begin{cases} w_{\frac{n+1}{2}} = 1 & \text{if } n \text{ is odd} \\ w_{\frac{n}{2}} = \frac{1}{2} \text{ and } w_{\frac{n}{2}+1} = \frac{1}{2} & \text{if } n \text{ is even} \\ w_i = 0 & \text{if } i \neq n \end{cases}$

Apart from these, other approaches [13, 14, 138, 170, 187, 188, 189, 196, 197, 199] for obtaining the OWA weights can be classified into two categories, namely: *argument-independent* and *argument-dependent*. As reflected by their respective names, the weights derived by the former are not related to the arguments being aggregated, while the latter determines the weights on the basis of the input arguments. In particular, the second category is considered in this thesis and several approaches in this group are reviewed.

2.5.1.1 DOWA Operator

The Dependent OWA (DOWA) operator [187] can relieve the influence of the unfair arguments on the aggregated result(s), where a normal distribution of argument values is assumed to determine their similarity degrees and, hence, the weights. In particular, a high weight is given to the argument whose value is close to the centre of all arguments (i.e., mean), whereas lower weights are assigned to those further away.

Let $(a_1, a_2, \ldots, a_n)$ be an argument vector, and e be the average value of this argument set: $e = \frac{1}{n}\sum_{i=1}^{n} a_i$. The similarity degree between any argument a_i and the average value e is calculated by

$$s(a_i, e) = 1 - \frac{|a_i - e|}{\sum_{j=1}^{n} |a_j - e|} \tag{2.85}$$

Note that if $\sum_{j=1}^{n} |a_j - e| = 0$, then $a_j - e = 0$, $j \in \{1, \ldots, n\}$. That is, all the values of the arguments are the same. In this case, $s(a_i, e) = 1$, $i \in \{1, \ldots, n\}$.

From this, an input vector $(a_1, a_2, \ldots, a_n)$ can be aggregated by the DOWA operator as follows:

$$\text{DOWA}(a_1, a_2, \ldots, a_n) = \sum_{j=1}^{n} w_j a_j \tag{2.86}$$

where the weight vector $W = (w_1, w_2, \ldots, w_n)^T$ is generated by

$$w_i = \frac{s(a_i, e)}{\sum_{j=1}^{n} s(a_j, e)}, \quad i \in \{1, \ldots, n\} \tag{2.87}$$

2.5.1.2 Clus-DOWA Operator

The cluster-based DOWA (Clus-DOWA) operator [13, 14] extends the DOWA operator and applies a distributed structure of data or data clusters in order to determine the weight vector. Those values very far from the group centre (i.e., mean) are not assigned with low weights, if they are seemingly indifferent to their local neighbours. An agglomerative hierarchical clustering technique [57] is then exploited to create the clustering structure for the studied values. In essence, the distance to the nearest cluster is employed to evaluate the *reliability* of each argument value and its assigned weight.

Let $(a_1, a_2, \ldots, a_n)$ be an argument vector. For each argument a_j, the concept of its reliability r_j is defined as its distance d_j to the nearest cluster recorded during a given clustering process, i.e.,

$$r_j = 1 - \frac{d_j}{\sum_{i=1}^{n} d_i} \tag{2.88}$$

Note that if $\sum_{i=1}^{n} d_i = 0$, then $d_i = 0$, $i \in \{1, \ldots, n\}$. This is a similar case to that mentioned previously, therefore $r_j = 1$, $j \in \{1, \ldots, n\}$.

From this, a specific and powerful OWA operator can be defined as follows. The Clus-DOWA operator is defined by

$$\text{Clus-DOWA}(a_1, a_2, \ldots, a_n) = \sum_{i=1}^{n} w_i a_i \tag{2.89}$$

where the weight vector is calculated from a computed vector of reliability measurement $(r_1, r_2, \ldots, r_n)$:

$$w_i = \frac{r_i}{\sum_{j=1}^{n} r_j}, \quad i \in \{1, \ldots, n\} \tag{2.90}$$

2.5.1.3 IOWA Operator

The induced OWA (IOWA) operator [199] takes as the argument pairs, called OWA pairs, in which one component is used to induce an ordering over the second components which are then aggregated. Central to this operator is the reordering of the arguments, based upon their values. That is, the weights rather than being associated with a specific argument, as in the case of the usual weighted average, are associated with a particular position in the ordering.

Let $(a_1, a_2, \ldots, a_n)$ be an argument vector. The ordering of the a_i $(i \in \{1, \ldots, n\})$ is induced by the so-called order inducing variables u_i $(i \in \{1, \ldots, n\})$, where u_i and a_i are the components of the OWA pairs $< u_i, a_i >$. The IOWA operator is defined as follows:

$$\text{IOWA}(< u_1, a_1 >, \ldots, < u_n, a_n >) = \sum_{j=1}^{n} w_j b_j \tag{2.91}$$

where $W = (w_1, w_2, \ldots, w_n)^T$ is a weight vector such that $w_i \in [0, 1]$, $\sum_{i=1}^{n} w_i = 1$, b_j is the a_i value of the OWA pair having the jth largest u_i $(i \in \{1, \ldots, n\})$, and u_i in $< u_i, a_i >$ is referred to as the order inducing variable and a_i as the argument variable.

2.5.2 Fuzzy Set Aggregation

Group decision making (GDM) involves the process of arriving at a judgement based upon the input and feedback of a group of individuals, which is at the same time beyond the competence of an individual. The group works cooperatively to achieve a satisfactory solution for all individuals concerned. As such, the solution is the one that is the most acceptable by the group of individuals as a whole [127].

In GDM problems, situations of partial agreement or even conflict amongst individuals may arise. Hence, finding a group consensus to represent a common opinion of the group is an important issue [74]. An appealing reason for using the information provided by several individuals when solving a problem is that a group-based approach may produce better solutions absorbing in different opinions. When facing such a decision problem, in order to avoid the impact of individual subjective judgement, choice and preference upon the final decision, the decision makers should gather all available relevant information. Such information is then aggregated so as to better form an impression of the problem and then make a decision.

Given the nature of GDM, subjectivity, imprecision and vagueness often appear in the assessment of the information to be aggregated. Thus, FS theory may play an important role in dealing with the problem of aggregation [55, 74]. As argued previously, much knowledge in the real-world is fuzzy rather than precise, and it is often the case that while real-world GDM problems can be handled easily by humans, they are often too difficult to be handled by machines. This observation has led to an increasing demand to improve machines' capability in handling fuzzy GDM problems, where decisions are automatically made in a fuzzy environment [20, 30, 31].

In general, a fuzzy GDM problem involves a finite number of alternatives and a finite set of experts whose opinions are concerned with imprecise data or information. That is, each expert may have a vague information about the performance of each alternative, and cannot estimate his/her preference with an exact numerical crisp or discrete value. Finding a solution to such a problem often needs to deal with linguistic assessments and natural language of the human expert, rather than exact numerical values. Each variable involved may therefore required to be assessed by means of linguistic terms or FSs [175]. A significant number of aggregation approaches based on FS theory have been proposed in order to address such problems [33, 74, 119, 120, 126, 171, 182]. Several representative methods are briefly outlined in the following subsections, which will be referred to in the subsequent development of the work reported in this thesis.

2.5.2.1 SAM

The similarity aggregation method (SAM) [74] aggregates the individual opinions that are subjectively estimated by experts and represented by trapezoidal fuzzy numbers. It first measures the degree of agreement between any two fuzzy opinions by using a function of pairwise similarity. For each fuzzy opinion, the degrees of agreement with respect to the other opinions are averaged. From this, the average degrees of agreement are normalised and combined linearly with the relative importance weights of experts to obtain the final composite weights for aggregating the individual expert opinions. However, this approach is not applicable to problems where the fuzzy opinions of experts do not overlap. In this case, the degree of similarity is zero. If all FSs representing opinions are disjoint, the aggregation process fails.

2.5.2.2 OAM

The optimal aggregation method (OAM) [120] aggregates the optimal consensus of expert opinions in the fuzzy GDM environment, where the importance of each expert is taken into consideration in the process of aggregation. This approach minimises the sum of weighted dissimilarities between the aggregated consensus and the individual opinions. One of the advantages of this approach is that it is valid even in the case when fuzzy opinions are disjoint. Also, it determines the weights using an optimisation model and is therefore optimal with respect to the criterion of the model. However, this approach is non-linear and computationally complex, which renders its application impractical in real-time fuzzy group decision analysis.

2.5.2.3 LSDM and DLSM

The least squares distance method (LSDM) and the defuzzification-based least squares method (DLSM) are proposed in [182] in order to overcome the drawbacks of the OAM. The former minimises the sum of squared distances from one weighted fuzzy opinion to another, and the latter minimises the sum of squared differences between the defuzzified values of any two weighted fuzzy opinions. One of the advantages of these two approaches is their simplicity, which is due to their closed-form expressions eliminate the need to perform the time-consuming iterative procedures. In addition, they can be utilised for aggregating interval numbers, triangular and trapezoidal fuzzy numbers, and even their combinations regardless of whether or not they overlap.

2.6 Summary

This chapter has introduced basic concepts of and recent developments in fuzzy interpolative reasoning, which supports inference with sparse fuzzy rule bases. Generally, the implementations of FRI can be categorised into two groups: one interpolating the consequence directly from a given observation, and the other following a two-step approach. The latter approach first generates an intermediate rule such that its antecedent part is as close to the given observation as possible, and then this intermediate rule is fired by the given observation through similarity-based fuzzy reasoning. It is this approach that the work to be developed in this thesis will follow.

The original KH approach and the T-FRI approach have been taken as representatives of the two groups in this chapter. The implementations of both approaches have been discussed, including the basic case, multiple antecedents case, and multiple rules case. Further, a review has been provided for the typical techniques that were developed in order to modify and improve the KH approach that may arrive a result which is not an FS.

In addition, to facilitate the establishment of a higher order framework for FRI in the next chapter, which can cope with more sophisticated uncertain information, underlying mathematical concepts such as RSs and type-2 FSs have also been introduced. Furthermore, as the foundations for extending the proposed framework, information aggregation techniques have been briefly reviewed, including the OWA-based aggregation and the similarity-based aggregation.

Chapter 3

Framework for Higher Order Representation and Interpolation

Fuzzy rule interpolation (FRI) forms an important approach for performing inference with systems comprising sparse rule bases. Even when a given observation has no overlap with the antecedent values of any existing rules, FRI may still derive a useful conclusion. However, little existing work on FRI can conjunctively handle more than one form of uncertainty in the rules. As argued previously, the difficulty in defining the required precise-valued MFs for the FSs that are used by conventional FRI techniques significantly restricts their application.

To overcome such difficulties, a novel framework is presented in this chapter for representing the knowledge involving higher order uncertainty and facilitating interpolation with such knowledge. It can handle both the first order and higher order types of uncertainty coherently. The proposed framework allows transformation-based rule interpolative techniques to be utilised in implementing a working higher order FRI system.

3.1 Basic Notions

The start point for the proposed framework is the requirement of being able to represent complicated uncertain knowledge in an effort to perform FRI. When exact membership values are no longer suitable for depicting the underlying uncertainty, it is desirable to utilise a certain higher order representation.

A *higher order* representation is a representation with the *first order* representation embedded within it. In this case, higher order representations are more expressive than the first order representation, whilst if the higher order knowledge degenerates to the first order, the computational mechanism that deals with higher order expressions is expected to naturally degenerate to the corresponding embedded first order calculus.

Practically, the notions of the lower bound and upper bound (of the uncertainty) are often designed for capturing and describing the ranges of uncertain knowledge. These two bounds consider the possible uncertain information and help construct different shaped uncertainty regions in the representation of the uncertainty.

Definition 3.1. Let X be the universe, a *higher order fuzzy set* (HOFS) $\tilde{A}$ can be represented by the *lower bound* $\tilde{A}^L$ and the *upper bound* $\tilde{A}^U$ such that

$$\tilde{A} =< x, [\mu^L_{\tilde{A}}(x), \mu^U_{\tilde{A}}(x)] >=< \tilde{A}^L, \tilde{A}^U >, \ \forall x \in X \tag{3.1}$$

where $0 \le \mu^L_{\tilde{A}}(x) \le \mu^U_{\tilde{A}}(x) \le 1$, and the lower and upper bounds are two conventional FSs, namely, two first order FSs.

Remark 3.1. The closer the shapes of $\tilde{A}^L$ and $\tilde{A}^U$ are, the less uncertain the information contained within $\tilde{A}$ is. When $\tilde{A}^L$ coincides with $\tilde{A}^U$, the HOFS degenerates to a conventional FS, i.e., $\mu^L_{\tilde{A}}(x) = \mu^U_{\tilde{A}}(x)$, $\forall x \in X$.

An important concept to introduce is the “less than” relation between two FSs [106]. An ordinary (type-1) set A_1 is said to be *less than* another ordinary FS A_2, denoted by $A_1 \prec A_2$, if $\forall \alpha \in (0,1]$, the following conditions hold:

$$\inf\{A_{1\alpha}\} < \inf\{A_{2\alpha}\}, \ \sup\{A_{1\alpha}\} < \sup\{A_{2\alpha}\} \tag{3.2}$$

where $A_{1\alpha}$ and $A_{2\alpha}$ are the α-cut sets of A_1 and A_2, respectively, $\inf\{A_{i\alpha}\}$ is the infimum of $A_{i\alpha}$, and $\sup\{A_{i\alpha}\}$ is the supremum of $A_{i\alpha}$, $i = 1, 2$.

Definition 3.2. An HOFS $\tilde{A}_1$ is said to be *less than* another HOFS $\tilde{A}_2$, denoted as $\tilde{A}_1 \tilde{\prec} \tilde{A}_2$, if and only if

$$\tilde{A}^L_1 \prec \tilde{A}^L_2, \ \tilde{A}^U_1 \prec \tilde{A}^U_2 \tag{3.3}$$

From this, the notion of neighbouring rules involving HOFSs can be defined.

Definition 3.3. Two higher order fuzzy rules

$$R_1: \text{If } x_1 \text{ is } \tilde{A}_{11},\ x_2 \text{ is } \tilde{A}_{12},\ \cdots,\ x_M \text{ is } \tilde{A}_{1M}, \text{ then } y \text{ is } \tilde{B}_1$$
$$R_2: \text{If } x_1 \text{ is } \tilde{A}_{21},\ x_2 \text{ is } \tilde{A}_{22},\ \cdots,\ x_M \text{ is } \tilde{A}_{2M}, \text{ then } y \text{ is } \tilde{B}_2$$

are said to be *neighbouring rules* if and only if: (1) $\tilde{A}_{1j} \precsim \tilde{A}_{2j}$ or $\tilde{A}_{2j} \precsim \tilde{A}_{1j}$, $j \in \{1, \cdots, M\}$ (where M is the number of antecedent variables in both rules); and (2) there is no individual rule "If x_1 is $\tilde{A}'_1$, x_2 is $\tilde{A}'_2$, $\cdots$, x_M is $\tilde{A}'_M$, then y is $\tilde{B}'$" such that $\tilde{A}_{1j} \precsim \tilde{A}'_j \precsim \tilde{A}_{2j}$ if $\tilde{A}_{1j} \precsim \tilde{A}_{2j}$, or $\tilde{A}_{2j} \precsim \tilde{A}'_j \precsim \tilde{A}_{1j}$ if $\tilde{A}_{2j} \precsim \tilde{A}_{1j}$, $j \in \{1, \cdots, M\}$.

Higher order FRI can then be achieved by extending the conventional FRI. In this case, the input and output of an interpolative process are HOFSs rather than conventional FSs.

Definition 3.4. Given a higher order fuzzy rule base and a higher order observation vector, *higher order FRI* is a process through which a conclusion from the given observation vector is obtained by interpolating the identified neighbouring rules which flank the observation that are taken from the rule base.

Note that in the above definition, two rules (e.g., the R_1 and R_2 given previously) are said to flank a given observation [106], say, $O = (\tilde{A}^*_1, \tilde{A}^*_2, \cdots, \tilde{A}^*_M)$, if $\tilde{A}_{1j} \precsim \tilde{A}^*_j \precsim \tilde{A}_{2j}$, or $\tilde{A}_{2j} \precsim \tilde{A}^*_j \precsim \tilde{A}_{1j}$, $j \in \{1, \cdots, M\}$.

3.2 Representative Values

In order to support the interpolation of rules involving HOFSs, the concept of *representative value* (Rep) is needed to be introduced. For simplicity, in this work, it is assumed that only polygonal HOFSs are considered; that is, both the lower and the upper bound are each represented by a polygonal-shaped first order FS. The Rep value captures important information such as the overall location of a (higher order) FS within the definition domain, and is computed and then utilised as the guide to perform subsequent inference during the interpolation process. The definition of Rep in HOFS follows the original definition in the existing T-FRI [76, 77], where for an ordinary FS A, the Rep(A) is calculated by

$$\text{Rep}(A) = \sum_{i=0}^{k-1} w_i a_i \tag{3.4}$$

with $A = (a_0, \cdots, a_{k-1})$ being a polygonal FS of k odd points, and w_i denoting the weight assigned to the point a_i.

Definition 3.5. Suppose that a polygonal HOFS $\tilde{A}$ is given, as shown in Figure 3.1, whose lower and upper bounds are: $\tilde{A} =< (\tilde{a}_0^L, \cdots, \tilde{a}_{l-1}^L; \tilde{H}_{\tilde{A}_1}^L, \cdots, \tilde{H}_{\tilde{A}_{l-2}}^L), (\tilde{a}_0^U, \cdots, \tilde{a}_{u-1}^U; \tilde{H}_{\tilde{A}_1}^U, \cdots, \tilde{H}_{\tilde{A}_{u-2}}^U) >$. The lower and upper Reps $\mathrm{Rep}(\tilde{A}^L)$ and $\mathrm{Rep}(\tilde{A}^U)$ of $\tilde{A}$ are defined by

$$\begin{cases} \mathrm{Rep}(\tilde{A}^L)_x = \sum_{i=0}^{l-1} w_i^L \tilde{a}_i^L \\ \mathrm{Rep}(\tilde{A}^L)_y = \sum_{i=1}^{l-2} w_i^L \tilde{H}_{\tilde{A}_i}^L \end{cases} \qquad \begin{cases} \mathrm{Rep}(\tilde{A}^U)_x = \sum_{j=0}^{u-1} w_j^U \tilde{a}_j^U \\ \mathrm{Rep}(\tilde{A}^U)_y = \sum_{j=1}^{u-2} w_j^U \tilde{H}_{\tilde{A}_j}^U \end{cases} \tag{3.5}$$

where w_v^V ($V \in \{L, U\}$, $v \in \{i, j\}$) is the weight assigned to the point $\tilde{a}_v^V$ and its corresponding membership value $\tilde{H}_{\tilde{A}_v}^V$, and x and y denote a certain variable dimension and the corresponding membership distribution, respectively.

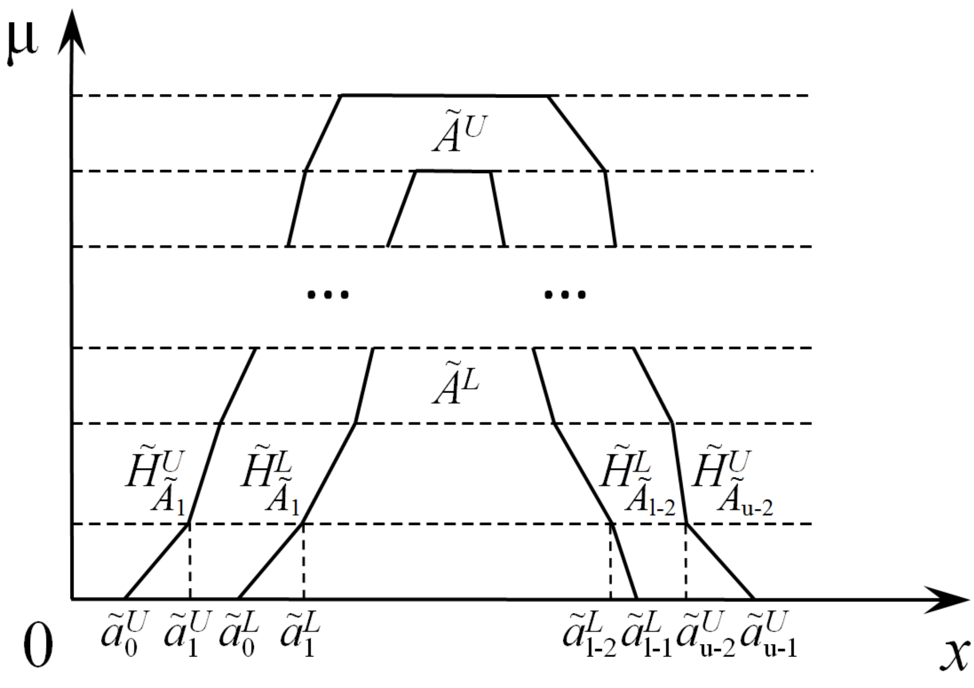


Figure 3.1: Lower bound $\tilde{A}^L$ and upper bound $\tilde{A}^U$ of a polygonal HOFS $\tilde{A}$

In general, specifying the weights is necessary for a given application. Different definitions can be adopted for deriving different Rep values. For instance, the simplest case is that all points take the same weight value, i.e., $w_i^L = 1/l$ and

$w_j^U = 1/u$. The centre of core can also be used to as an alternative. In this case, the Rep may be solely determined by those points with a membership value of 1: $\text{Rep}(\tilde{A}^U)_x = \frac{1}{2}(\tilde{a}^U_{\lceil (u/2) \rceil - 1} + \tilde{a}^U_{u - \lceil (u/2) \rceil})$ and $\text{Rep}(\tilde{A}^U)_y = \frac{1}{2}(\tilde{H}^U_{\tilde{A}_{\lceil (u/2) \rceil - 1}} + \tilde{H}^U_{\tilde{A}_{u - \lceil (u/2) \rceil}})$. The lower Reps are omitted here, which can be calculated in a similar way involving those points of the maximum membership value. Other alternative definitions can be found in [77].

Remark 3.2. In the existing T-FRI $\text{Rep}(A)_y$ of a given conventional FS A is a constant, only the x value is therefore considered. However, this is no longer the case in this framework due to the introduction of higher order uncertainty, both x and y dimensions need to be considered. The calculation of $\text{Rep}(\tilde{A})_y$ follows that used to calculate $\text{Rep}(\tilde{A})_x$ to maintain consistency.

In order to distinguish amongst different HOFS shapes, the *shape diversity factor* f is herein introduced. This work follows the conventional definition of statistical standard deviation (although this may be defined differently if desired for a particular implementation).

Definition 3.6. The lower and upper shape diversity factors $f^L_{\tilde{A}}$ and $f^U_{\tilde{A}}$ are defined by

$$\begin{cases} f^L_{\tilde{A}} = \sqrt{\dfrac{\sum\limits_{i=0}^{l-1}(\tilde{a}^L_i - \text{Rep}(\tilde{A}^L)_x)^2}{l}} \\ f^U_{\tilde{A}} = \sqrt{\dfrac{\sum\limits_{j=0}^{u-1}(\tilde{a}^U_j - \text{Rep}(\tilde{A}^U)_x)^2}{u}} \end{cases} \tag{3.6}$$

Remark 3.3. A small shape diversity factor implies that the odd points of $\tilde{A}^L$ ($\tilde{A}^U$) tend to be close to those of the lower (upper) Rep. That is, the smaller the shape diversity factor, the smaller the area of the lower (upper) bound.

Extending T-FRI to FRI involving HOFSs, a single overall Rep of a given HOFS is required. For this, the *weight factor* w of the lower (upper) bound is first introduced below, which reflects the relative contribution of the lower (upper) shape diversity in depicting the underlying HOFS. The introduction of these lower and upper shape diversity factors helps minimise the opportunity of having the same Rep value from the use of HOFSs of different shapes.

Definition 3.7. The lower and upper weight factors $w^L_{\tilde{A}}$ and $w^U_{\tilde{A}}$ are defined as the weights of the shape diversity factors, in terms of the areas of the lower and upper bounds, such that:

$$w^V_{\tilde{A}} = \frac{f^V_{\tilde{A}}}{f^L_{\tilde{A}} + f^U_{\tilde{A}}}, \quad V = L, U \tag{3.7}$$

Remark 3.4. In general, $f^L_{\tilde{A}} + f^U_{\tilde{A}} \neq 0$. If however, $f^L_{\tilde{A}} + f^U_{\tilde{A}} = 0$, i.e., $f^L_{\tilde{A}} = 0$ and $f^U_{\tilde{A}} = 0$, the HOFS degenerates to a singleton value, $w^L_{\tilde{A}} = w^U_{\tilde{A}} = 1/2$.

Definition 3.8. The overall Rep of a given HOFS $\tilde{A}$ is defined by

$$\mathrm{Rep}(\tilde{A}) = \sum_{V \in \{L,U\}} (w^V_{\tilde{A}} \sum_{e \in \{x,y\}} \mathrm{Rep}(\tilde{A}^V)_e) \tag{3.8}$$

where $w^V_{\tilde{A}}$ is the weight assigned to $\mathrm{Rep}(\tilde{A}^V)$ of $\tilde{A}^V$, $V \in \{L, U\}$.

As with the first order methods, in general, multiple rules with multiple antecedents need to be taken into consideration in order to obtain an interpolated conclusion. For this, the first step that needs to be considered is to choose the closest N ($N \geq 2$) rules from the rule base with respect to the given observation. A *distance* measure is thus utilised to measure the proximity of the rules by exploiting such Rep values that capture specific information embedded in HOFSs.

3.3 Selection of Closest N Rules

Without losing generality, suppose that there are n higher order fuzzy rules in a higher order fuzzy rule base. A rule R_i, an observation O and the conclusion C are represented by the following, respectively:

$$R_i : \text{If } x_1 \text{ is } \tilde{A}_{i1}, \cdots, x_j \text{ is } \tilde{A}_{ij}, \cdots, x_M \text{ is } \tilde{A}_{iM}, \text{ then } y \text{ is } \tilde{B}_i$$
$$O : x_1 \text{ is } \tilde{A}^*_1, \cdots, x_j \text{ is } \tilde{A}^*_j, \cdots, x_M \text{ is } \tilde{A}^*_M$$
$$C : y \text{ is } \tilde{B}^*$$

where $\tilde{A}_{ij}$ denotes the jth antecedent HOFS of R_i, $\tilde{A}^*_j$ is the observation of the variable x_j, $\tilde{B}^*$ is the desired interpolated conclusion, and $\tilde{B}_i$ denotes the consequent HOFS of R_i with $j \in \{1, \cdots, M\}$, with M being the number of antecedent variables.

Definition 3.9. The distance d_{ij} between the pair of $\tilde{A}_{ij}$ and $\tilde{A}_j^*$ is defined as follows:

$$d_{ij} = d(\tilde{A}_{ij}, \tilde{A}_j^*) = d(\mathrm{Rep}(\tilde{A}_{ij}), \mathrm{Rep}(\tilde{A}_j^*)) \tag{3.9}$$

where $d(.,.)$ is herein computed using the Euclidean distance metric (though any other distance metric may be used as an alternative).

Definition 3.10. The distance d_i between the rule R_i and the observation O is deemed to be the average of the distances between the HOFSs of each rule antecedent and the corresponding variable in O:

$$d_i = \sqrt{\sum_{j=1}^{M} {d'_{ij}}^2}, \quad d'_{ij} = \frac{d_{ij}}{\max_j - \min_j} \tag{3.10}$$

where $\max_j$ and $\min_j$ are the maximum and minimum value in the domain of the variable x_j, $j \in \{1, \ldots, M\}$. Each distance measure d_{ij} is normalised into the range $[0, 1]$, denoted by d'_{ij}, to ensure the resulting distances to be compatible with each other over different domains. Note that if $\max_j - \min_j = 0$, then $\max_j = \min_j$. That is, $\tilde{A}_j^*$ is identical with $\tilde{A}_{ij}$, $j \in \{1, \ldots, M\}$. In this case, $d'_{ij} = 1$.

Given the above definition, the distances between a given observation and all rules in the rule base can be calculated. The N rules which have minimal distances are chosen as the closest N rules with respect to the given observation. The choice of a larger N will help consider a wider range of neighbouring rules in performing interpolation, thereby more likely to result in global results but requiring significantly more computation. On the contrary, the choice of a relatively smaller N will tend to considering only neighbouring rules and hence involving less computation time. Since FRI is in general used to derive an approximate result in the first place, in practical application, N can be chosen to be 2. This is the case for conventional rule interpolation also. However, in the following theoretical development to maintain generality, the number of closest rules is set to N ($N \geq 2$) unless otherwise stated.

3.4 Construction of Intermediate Rule

As with a number of first order FRI approaches, higher order FRI is in this work developed following the principle of analogical reasoning [15]. First, an artificially created intermediate rule is interpolated such that the antecedent of the intermediate rule is as "close" to the given observation as possible. Then, a conclusion is worked out from the given observation by firing this generated intermediate rule through a certain analogical reasoning mechanism.

Definition 3.11. Suppose that N closest rules are chosen with respect to a given observation. These rules are represented as R_i, $i \in \{1,\ldots,N\}$, each having M antecedent variables $\tilde{A}_{ij}$, $j \in \{1,\ldots,M\}$ and are used to derive the intermediate rule. Let $w_{\tilde{A}_{ij}}$ denote the weight to which the jth antecedent of the ith closest rule contributes to the emerging intermediate rule, which is defined as the reciprocal of the corresponding distance measure:

$$w_{\tilde{A}_{ij}} = \frac{1}{d_{ij}} = \frac{1}{d(\tilde{A}_{ij},\tilde{A}_j^*)} \tag{3.11}$$

where $\tilde{A}_j^*$ denotes the observed HOFS of antecedent variable j. The normalised weight $w'_{\tilde{A}_{ij}}$ is then defined by

$$w'_{\tilde{A}_{ij}} = \frac{w_{\tilde{A}_{ij}}}{\sum_{i=1}^{N} w_{\tilde{A}_{ij}}} \tag{3.12}$$

Remark 3.5. This definition reflects the intuition that the larger the distance is, the less relevant the corresponding attribute is to the observation. In general, $d_{ij} \neq 0$. If however, $d_{ij} = 0$, then $\text{Rep}(\tilde{A}_{ij}) = \text{Rep}(\tilde{A}_j^*)$. In this case, the antecedent of the observation is considered to be identical to the corresponding antecedent of the rule R_i, in terms of their Rep values. Thus, $w_{\tilde{A}_{ij}}$ is set to 1 for the identical cases with the rest set to 0.

The antecedent $\tilde{A}_j^{IFT}$ of the intermediate rule is constructed from the antecedents of the identified closest rules. A process *shift* is then utilised to modify $\tilde{A}_j^{IFT}$ so that the antecedent of the intermediate rule will have the same Rep as $\tilde{A}_j^*$:

$$\tilde{A}'_j = \tilde{A}_j^{IFT} + \delta_{\tilde{A}_j}(\max_j - \min_j), \quad \tilde{A}_j^{IFT} = \sum_{i=1}^{N} w'_{\tilde{A}_{ij}} \tilde{A}_{ij} \tag{3.13}$$

where $\delta_{\tilde{A}_j}$ is a constant defined by

$$\delta_{\tilde{A}_j} = \frac{\text{Rep}(\tilde{A}_j^*) - \text{Rep}(\tilde{A}_j^{IFT})}{\max_j - \min_j} \tag{3.14}$$

Note that if $\max_j - \min_j = 0$, then $\max_j = \min_j$. That is, $\tilde{A}_j^*$ is identical with $\tilde{A}_j^{IFT}$, $j \in \{1,\ldots,M\}$. In this case, $\delta_{\tilde{A}_j} = 1$.

The consequence of the intermediate rule $\tilde{B}'$ is calculated by analogy to the computation of the antecedent, such that:

$$\tilde{B}' = \tilde{B}^{IFT} + \delta_{\tilde{B}}(\max - \min), \quad \tilde{B}^{IFT} = \sum_{i=1}^{N} w'_{\tilde{B}_i} \tilde{B}_i \tag{3.15}$$

where $\tilde{B}^{IFT}$ is the consequence of the intermediate fuzzy rule, max and min are the maximum and minimum values within the domain of the consequent variable, $w'_{\tilde{B}_i}$ and $\delta_{\tilde{B}}$ are the means of $w'_{\tilde{A}_{ij}}$ and $\delta_{\tilde{A}_j}$, $i \in \{1, \ldots, N\}$, $j \in \{1, \ldots, M\}$, respectively, which are defined by

$$w'_{\tilde{B}_i} = \frac{1}{M} \sum_{j=1}^{M} w'_{\tilde{A}_{ij}}, \quad \delta_{\tilde{B}} = \frac{1}{M} \sum_{j=1}^{M} \delta_{\tilde{A}_j} \tag{3.16}$$

3.5 Interpolation through Similarity-constrained Transformations

The aforementioned artificially constructed intermediate rule is derived from the chosen closest rules with respect to an observation. It can be used to perform inference without further reference to its originals. Suppose that a certain degree of similarity between the antecedent part of this rule and the observation is established, it is intuitive to require that its consequent part and the eventual conclusion to be drawn attain the same similarity degree. That is, for an intermediate rule: "If x_1 is $\tilde{A}'_1$, $\cdots$, x_j is $\tilde{A}'_j$, $\cdots$, x_M is $\tilde{A}'_M$, then y is $\tilde{B}'$", and a given observation $O = (\tilde{A}^*_1, \cdots, \tilde{A}^*_j, \cdots, \tilde{A}^*_M)$, the shape distinguishability between $\tilde{B}'$ and the interpolated consequence $\tilde{B}^*$ is analogous to the shape distinguishabilities between $\tilde{A}'_j$ and $\tilde{A}^*_j$, $j = 1, 2, \cdots, M$. In order to ensure this, the following three transformations are designed.

Note that all three transformations are separately implemented on each dimension and separately calculated on each of the lower and upper bound. However, the underlying computational mechanisms are identical. For presentational simplicity, the description of these transformations is given without the subscript j and the superscript L or U.

3.5.1 Scale Transformation

Consider the lower (upper) bound of $\tilde{A}'$ and that of $\tilde{A}^*$, respectively represented as $\tilde{A}' = (\tilde{a}'_0, \cdots, \tilde{a}'_{k-1}; \tilde{H}'_{\tilde{A}_1}, \cdots, \tilde{H}'_{\tilde{A}_{k-2}})$ and $\tilde{A}^* = (\tilde{a}^*_0, \cdots, \tilde{a}^*_{k-1}; \tilde{H}^*_{\tilde{A}_1}, \cdots, \tilde{H}^*_{\tilde{A}_{k-2}})$. The following parameters, termed the scale rates s_p $(p = 0, \cdots, \lfloor (k/2) \rfloor - 1)$ rescale the pth support of $\tilde{A}'$ to approximate that of $\tilde{A}^*$:

$$s_p = \frac{\tilde{a}^*_{k-p-1} - \tilde{a}^*_p}{\tilde{a}'_{k-p-1} - \tilde{a}'_p} \tag{3.17}$$

From these scale rates, the following scale ratios $\mathbb{S}_q$ $(q = 1, \cdots, \lfloor (k/2) \rfloor - 1)$ modify the rescaled qth support of $\tilde{A}'$ to further approximate that of $\tilde{A}^*$ such that the resulting HOFS $\tilde{A}'$ is of the same scale as that of $\tilde{A}^*$:

$$\mathbb{S}_q = \begin{cases} \dfrac{\frac{\tilde{a}^*_{k-q-1} - \tilde{a}^*_q}{\tilde{a}^*_{k-q} - \tilde{a}^*_{q-1}} - \frac{\tilde{a}'_{k-q-1} - \tilde{a}'_q}{\tilde{a}'_{k-q} - \tilde{a}'_{q-1}}}{1 - \frac{\tilde{a}'_{k-q-1} - \tilde{a}'_q}{\tilde{a}'_{k-q} - \tilde{a}'_{q-1}}} & \text{if } s_q \geq s_{q-1} \\ \dfrac{\frac{\tilde{a}^*_{k-q-1} - \tilde{a}^*_q}{\tilde{a}^*_{k-q} - \tilde{a}^*_{q-1}} - \frac{\tilde{a}'_{k-q-1} - \tilde{a}'_q}{\tilde{a}'_{k-q} - \tilde{a}'_{q-1}}}{\frac{\tilde{a}'_{k-q-1} - \tilde{a}'_q}{\tilde{a}'_{k-q} - \tilde{a}'_{q-1}}} & \text{if } s_{q-1} > s_q \end{cases} \tag{3.18}$$

From this, by imposing the required similarities, the corresponding scale rates s'_p that will help rescale the pth support of $\tilde{B}'$ into the emerging $\tilde{B}^*$ can be obtained such that

$$s'_p = \begin{cases} s_p & \text{if } p = 0 \\ \dfrac{s'_{p-1}(s_p - s_{p-1})(\frac{\tilde{b}'_{k-p} - \tilde{b}'_{p-1}}{\tilde{b}'_{k-p-1} - \tilde{b}'_p} - 1)}{s_{p-1}(\frac{\tilde{a}'_{k-p} - \tilde{a}'_{p-1}}{\tilde{a}'_{k-p-1} - \tilde{a}'_p} - 1)} + s'_{p-1} & \text{if } s_p \geq s_{p-1},\ p > 0 \\ \dfrac{s'_{p-1} s_p}{s_{p-1}} & \text{if } s_{p-1} > s_p,\ p > 0 \end{cases} \tag{3.19}$$

The above shows only the situation where one antecedent variable is considered (for either a lower bound or an upper bound). In general, for each antecedent variable j and each bound V, $V \in \{L, U\}$, such a scale transformation is repeatedly applied to transform $\tilde{A}^{'V}_j$ to the intermediate terms $\tilde{A}^{''V}_j$ with s^V_{jp} and $\mathbb{S}^V_{jq}$. $\tilde{B}^{''V}$ is then generated from $\tilde{B}^{'V}$ using the aggregated $s^V_{\tilde{B}_p}$ and $\mathbb{S}^V_{\tilde{B}_q}$, where $s^V_{\tilde{B}_p} = \frac{1}{M}\sum_{j=1}^M s^{'V}_{jp}$ and $\mathbb{S}^V_{\tilde{B}_q} = \frac{1}{M}\sum_{j=1}^M \mathbb{S}^{'V}_{jq}$.

3.5.2 Move Transformation

The move ratios $\mathbb{M}_r$ $(r = 0, \cdots, \lceil (k/2) \rceil - 2)$ shift the locations of supports of $\tilde{A}^{(r-1)}$ to that of $\tilde{A}^{*}$ (where $\tilde{A}^{(r-1)}$ is the term obtained after the (r-1)th sub-move with initialisation $\tilde{A}^{-1} = \tilde{A}''$):

$$\mathbb{M}_r = \begin{cases} \dfrac{\tilde{a}^{*}_r - \tilde{a}^{(r-1)}_r}{\min\{\frac{\tilde{a}^{(r-1)}_r + \cdots + \tilde{a}^{(r-1)}_{\lceil (k/2) \rceil - 1}}{\lceil (k/2) \rceil - r} - \tilde{a}^{(r-1)}_r, \tilde{a}^{(r-1)}_{k-r} - \tilde{a}^{(r-1)}_{k-r-1}\}} & \text{if } \tilde{a}^{*}_r \geq \tilde{a}^{(r-1)}_r \\ \dfrac{\tilde{a}^{*}_r - \tilde{a}^{(r-1)}_r}{\min\{\tilde{a}^{(r-1)}_{k-r-1} - \frac{\tilde{a}^{(r-1)}_{k-\lceil (k/2) \rceil} + \cdots + \tilde{a}^{(r-1)}_{k-r-1}}{\lceil (k/2) \rceil - r}, \tilde{a}^{(r-1)}_r - \tilde{a}^{(r-1)}_{r-1}\}} & \text{if } \tilde{a}^{(r-1)}_r > \tilde{a}^{*}_r \end{cases} \tag{3.20}$$

where $\tilde{a}^{(r-1)}_r$ is the $\tilde{a}''_r$'s new position after the (r-1)th sub-move. Initially, when $r = 0$, $\tilde{a}^{(-1)}_0 = \tilde{a}''_0$, $\tilde{a}^{(r-1)}_{k-r} - \tilde{a}^{(r-1)}_{k-r-1}$ and $\tilde{a}^{(r-1)}_r - \tilde{a}^{(r-1)}_{r-1}$ are not included into the calculation of $\min\{.,.\}$.

In general, for each antecedent variable j and each bound V, $V \in \{L, U\}$, this move transformation is repeatedly applied to obtain $\tilde{A}^{(r)V}_j = \{\tilde{a}^{(r)V}_{j0}, \cdots, \tilde{a}^{(r)V}_{j(k-1)}\}$ from $\tilde{A}^{(r-1)V}_j$ using $\mathbb{M}^V_{jr}$. $\tilde{B}^{(r)V} = \{\tilde{b}^{(r)V}_0, \cdots, \tilde{b}^{(r)V}_{k-1}\}$ is then obtained from $\tilde{B}^{(r-1)V}$ using the aggregated $\mathbb{M}^V_{\tilde{B}_r}$, where $\mathbb{M}^V_{\tilde{B}_r} = \frac{1}{M}\sum_{j=1}^{M} \mathbb{M}^V_{jr}$, resulting in $\tilde{A}^{(\lceil (k/2) \rceil - 2)V}_j = \tilde{A}^{*V}_j$ and $\tilde{B}^{(\lceil (k/2) \rceil - 2)V} = \tilde{B}^{*V}$.

3.5.3 Height Transformation

Due to the higher order uncertainty, the height rates h_o $(o = 1, \cdots, k-2)$ are utilised to adjust the heights $\tilde{H}'^{L}_{\tilde{A}_o}$ of $\tilde{A}'^{L}$ to the heights $\tilde{H}^{*L}_{\tilde{A}_o}$ of $\tilde{A}^{*L}$:

$$h_o = \frac{\tilde{H}^{*L}_{\tilde{A}_o}}{\tilde{H}'^{L}_{\tilde{A}_o}} \tag{3.21}$$

where $0 < \tilde{H}^{*L}_{\tilde{A}_o} \leq \tilde{H}^{*U}_{\tilde{A}_o} = 1$ and $0 < \tilde{H}'^{L}_{\tilde{A}_o} \leq \tilde{H}'^{U}_{\tilde{A}_o} = 1$. This constraint applies to the interpolated conclusion as well. That is, if the height of $\tilde{B}^{*L}$ is greater than the height of $\tilde{B}^{*U}$ after the height transformation, then $\tilde{H}^{*L}_{\tilde{B}_o} = \tilde{H}^{*U}_{\tilde{B}_o}$.

In general, for each antecedent variable j and each bound V, $V \in \{L, U\}$, this height transformation is repeatedly applied to transform the heights of $\tilde{A}'^{L}_j$ to the heights of $\tilde{A}^{*L}_j$ with h_{jo}. The height of the interpolated conclusion is then obtained using the aggregated $h_{\tilde{B}_o}$, where $h_{\tilde{B}_o} = \frac{1}{M}\sum_{j=1}^{M} h_{jo}$.

Remark 3.6. Scale transformation scales $\tilde{A}'_j$ up or down to $\tilde{A}''_j$ retaining the ratios between left and right slopes, but having different supports length. The closer the scale ratios to 0, the more similar $\tilde{A}'_j$ and $\tilde{A}''_j$. Move transformation shifts $\tilde{A}''_j$ to $\tilde{A}^*_j$ which has the same support length, but having different locations for the supports. The closer the move ratios to 0, the more similar $\tilde{A}''_j$ and $\tilde{A}^*_j$. Height transformation adjusts the height of $\tilde{A}'_j$ to the height of $\tilde{A}^*_j$ while the characteristics remain the same. The closer the height rates to 1, the more similar $\tilde{A}'_j$ and $\tilde{A}^*_j$.

Integrally, scale, move and height transformations guarantee that the transferred sets have the same type of shapes as that of the original. That is, these three transformations allow the similarity degree between $\tilde{B}'$ and $\tilde{B}^*$ to be measured by those between $\tilde{A}'_j$ and $\tilde{A}^*_j$.

3.6 Summary

In this chapter, a novel framework that consists of higher order knowledge representation and higher order rule interpolation has been presented. The proposed framework is on the basis of the transformation-based interpolative technique. It extends the application of the existing T-FRI to higher order environment, offering greater flexibility in handling different types of uncertainty that may be present in sparse rule bases and observations. Instead of addressing just the first order uncertainty like conventional FRI methods, the proposed framework can handle both the first order and higher order uncertain information coherently.

This chapter has presented a generic specification for higher order FRI in which the concept of HOFSs and the algorithm for higher order interpolation have been discussed. In particular, the algorithm works by first using the lower and upper Reps to approximate the lower and upper bounds of an HOFS, and then deriving an intermediate rule using the proportional value which is calculated by the Reps. Next, scale, move and height transformations are utilised in transformation-based interpolation to preserve the similarity degree between the observation and the antecedent(s) of the artificially created intermediate rule. Finally, the interpolated conclusion is computed by applying transformation functions to the consequence of the intermediate rule with the same similarity degree.

The above framework is proposed to allow the representation and application of higher order uncertainty knowledge for FRI. Different approaches can be implemented in this framework. The following two chapters present such specifications, one using RF sets (Chapter 4) and the other using type-2 FSs (Chapter 5).

Chapter 4

Implementing Framework with Rough-Fuzzy Sets

ROUGH set (RS) theory is a useful tool to deal with incomplete knowledge by the introduction of the concepts of lower and upper approximations. This chapter introduces a new extension to RS. Based on this, a rough-fuzzy (RF) approach to FRI is presented to demonstrate the flexibility of the previously proposed framework, by exploiting the concept of RF sets and generalising the T-FRI techniques. In particular, a refinement procedure is described in order to ensure intuitive interpolated conclusions. A proof is also provided to verify that the RF approach is indeed compatible with the original T-FRI.

4.1 Rough-Fuzzy Sets

The concept of RSs [151] was originally proposed as a mathematical tool to deal with incomplete or imperfect data and knowledge in information systems. An RS is itself an approximation of a vague concept by a pair of precise sets, called lower and upper approximations [150]. The lower approximation (LA) contains all of those objects which *definitely* belong to a concept, and the upper approximation (UA) contains all of those objects which *possibly* belong to the concept. RSs characterise the roughness of a set using these two approximations [5].

Inspired by this observation, it is useful to integrate rule interpolation with the RF concept in order to deal with higher order uncertainty. Such an implementation of the preceding framework is proposed here. It modifies the underlying FRI technique to ensure intuitive interpolated conclusions. In particular, this work facilitates the representation of uncertain fuzzy set (FS) membership functions (MFs) with RF approximations, thereby improving the flexibility of rule interpolation in dealing with higher order uncertainty in fuzziness.

Definition 4.1. With any $P \subseteq \mathbb{A}$, an alternative equivalence relation $IND(P)$ to the traditional one of Equation (2.75) can be defined by

$$IND(P) = \{(x, y) \in \mathbb{U}^2 \mid \forall F_g \in P,\ F_g(x) \in C_z,\ F_g(y) \in C_z\} \tag{4.1}$$

where F_g, $g \in \{1, \ldots, G\}$, are FSs that jointly define a particular concept C_z, where C_z, $z \in \{1, \ldots, Z\}$, is a concept in X, i.e., $X = \{C_1, C_2, \ldots, C_Z\}$, $X \subseteq \mathbb{U}$.

Equation (4.1) expresses the equivalence relation between the memberships of x and y to different FSs of given concept. Using this equivalence relation, the lower and upper approximations for each C_z in X can be redefined as follows.

Definition 4.2. Let $IND(P)$ be an equivalence relation on $\mathbb{U}$ and F_g, $g \in \{1, \ldots, G\}$, be FSs in C_z ($C_z \in X$), the lower and upper approximations are a pair of FSs with MFs defined by the following, respectively:

$$\begin{aligned} \mu_{\underline{P}C_z}(x \in [x]_P) &= \inf\{\mu_{F_g}(x), g \in \{1, \ldots, G\} | x \in [x]_P\} \\ \mu_{\overline{P}C_z}(x \in [x]_P) &= \sup\{\mu_{F_g}(x), g \in \{1, \ldots, G\} | x \in [x]_P\} \end{aligned} \tag{4.2}$$

The tuple $< \underline{P}X, \overline{P}X >$ is called an RF set (which differs from the alternative use of this term in the literature [5] due to parallel development of these related but different concepts).

Reconsider the situation shown in Chapter 1, where different people may interpret the same concept differently. As reflected in Figure 1.1, it is difficult to describe this situation using conventional FSs. However, the newly defined RF sets can be adopted to represent this uncertain concept by exploiting the two approximations. The LA indicates the intersection amongst regions that are agreed by individuals, while the UA indicates the union of the regions that are given by at least one person, as shown in Figure 4.1. RF sets therefore utilise LAs and UAs to express the different types of uncertainty involved in defining fuzzy memberships.

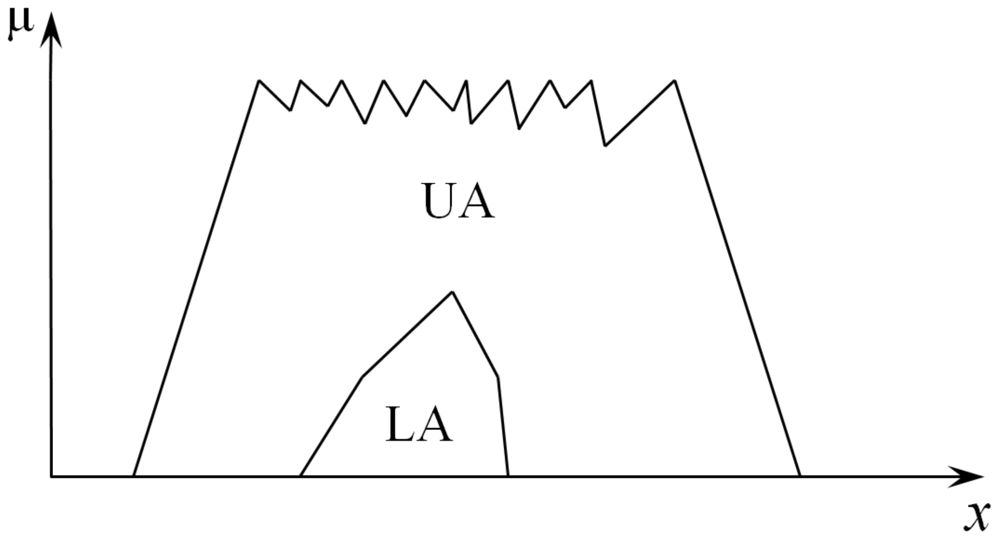


Figure 4.1: An RF set corresponding to the situation depicted by Figure 1.1

4.2 Rough-Fuzzy Implementation of the Framework

Using RF sets, the procedure of the proposed framework can be directly implemented. Proofs of the resulting computation methods are omitted here to save space. The algorithm for deriving the interpolated conclusion with multiple multi-antecedent rules is outlined below. Suppose that there are an RF rule base and an RF observation, the inference model for RF implementation can be represented by

$$\begin{aligned} R_1 &: \text{If } x_1 \text{ is } \tilde{A}_{11} \text{ and } x_2 \text{ is } \tilde{A}_{12}, \text{ then } y \text{ is } \tilde{B}_1 \\ R_2 &: \text{If } x_1 \text{ is } \tilde{A}_{21} \text{ and } x_2 \text{ is } \tilde{A}_{22}, \text{ then } y \text{ is } \tilde{B}_2 \\ O &: x_1 \text{ is } \tilde{A}^*_1 \text{ and } x_2 \text{ is } \tilde{A}^*_2 \end{aligned}$$

where for computational simplicity $\tilde{A}_{11}, \tilde{A}_{12}, \tilde{A}_{21}, \tilde{A}_{22}, \tilde{A}^*_1, \tilde{A}^*_2, \tilde{B}_1$ and $\tilde{B}_2$ are assumed to be trapezoidal RF sets, $\tilde{A}_{11} \wedge \tilde{A}_{12} \Rightarrow \tilde{B}_1$ and $\tilde{A}_{21} \wedge \tilde{A}_{22} \Rightarrow \tilde{B}_2$ are two adjacent and disjoint RF rules with each having two antecedent variables, as shown in Figure 4.2.

4.2.1 Calculating Representative Values

The lower and upper Reps, $\text{Rep}(\tilde{A}^{*V}_j)_x$ and $\text{Rep}(\tilde{A}^{*V}_j)_y$, are calculated first by

$$\begin{cases} \text{Rep}(\tilde{A}^{*V}_j)_x = \dfrac{1}{4}(\tilde{a}^{*V}_{j0} + \tilde{a}^{*V}_{j1} + \tilde{a}^{*V}_{j2} + \tilde{a}^{*V}_{j3}) \\ \text{Rep}(\tilde{A}^{*V}_j)_y = \dfrac{1}{4}(\tilde{H}^V_{\tilde{A}_{j1}} + \tilde{H}^V_{\tilde{A}_{j2}}) \end{cases} \tag{4.3}$$

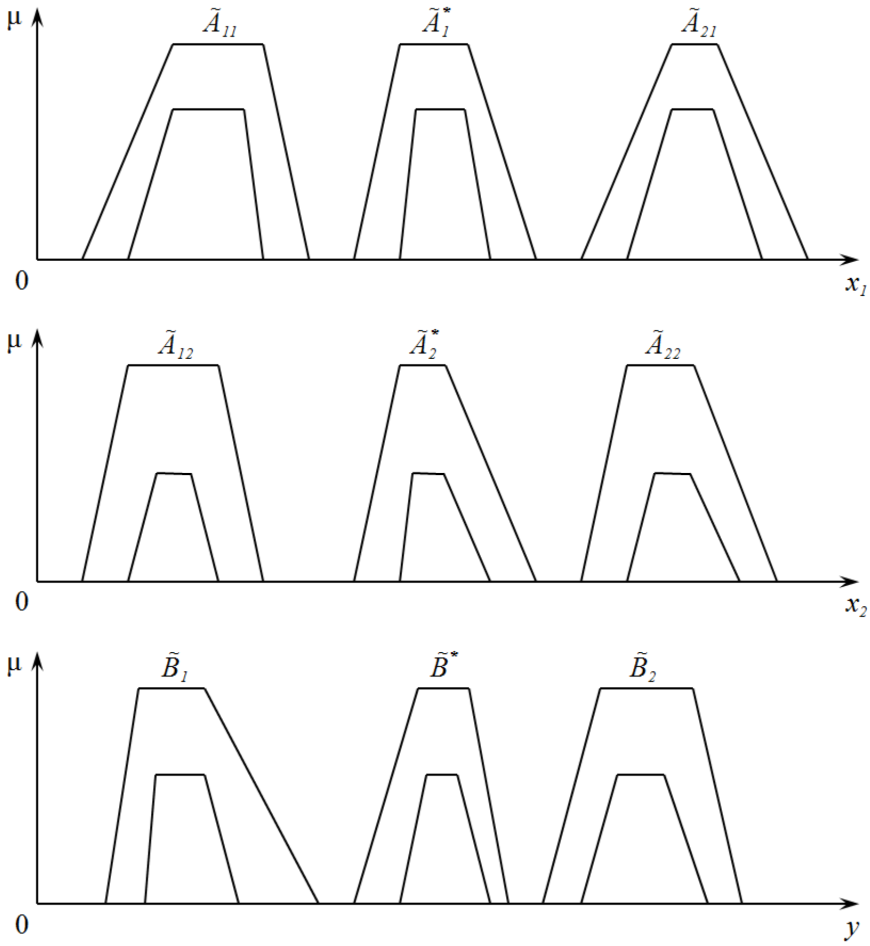


Figure 4.2: RF implementation with trapezoidal RF sets

where $j = 1, 2$ and $V = L, U$. As a special case of Equation (3.5), for simplicity, the weights assigned to points are herein determined by the arithmetic average. The shape diversity factors $f_{\tilde{A}_j}^{*V}$ and weight factors $w_{\tilde{A}_j}^{*V}$ are respectively computed by

$$f_{\tilde{A}_j^*}^{V} = \sqrt{\frac{\sum_{k=0}^{3}(\tilde{a}_{jk}^{*V} - \mathrm{Rep}(\tilde{A}_j^{*V})_x)^2}{4}} \tag{4.4}$$

$$w_{\tilde{A}_j}^{*V} = \frac{f_{\tilde{A}_j}^{*V}}{f_{\tilde{A}_j}^{*L} + f_{\tilde{A}_j}^{*U}} \tag{4.5}$$

where $j = 1, 2$ and $V = L, U$. The overall Reps $\mathrm{Rep}(\tilde{A}_j^*)$ are then obtained with

$$\mathrm{Rep}(\tilde{A}_j^*) = w_{\tilde{A}_j}^{*L}(\mathrm{Rep}(\tilde{A}_j^{*L})_x + \mathrm{Rep}(\tilde{A}_j^{*L})_y) + w_{\tilde{A}_j}^{*U}(\mathrm{Rep}(\tilde{A}_j^{*U})_x + \mathrm{Rep}(\tilde{A}_j^{*U})_y) \tag{4.6}$$

where $j = 1, 2$. The calculations for $\tilde{A}_{ij}$ ($i = 1, 2$ and $j = 1, 2$) follow the same procedure.

4.2.2 Choosing Two Closest Rules

The distances between a given observation O and two rules, say, R_1 and R_2 in the rule base are calculated using Equation (3.10) exactly. Here, two rules are chosen as the closest rules to perform interpolation, again for computational simplicity (although in general, $N, N \geq 2$, rules may be used).

4.2.3 Constructing Intermediate Rule

This step is exactly the same as given in the opposite number in the general framework. That is, the weight $w_{\tilde{A}_{ij}}$ of the jth antecedent of the ith chosen rule is computed by Equation (3.11). Its normalised weight $w'_{\tilde{A}_{ij}}$, which is calculated by Equation (3.12), together with the parameter $\delta_{\tilde{A}_j}$, which is calculated by Equation (3.14), is used in Equation (3.13) to obtain the antecedent of the intermediate rule $\tilde{A}'_j$ for each antecedent dimension x_j, $i = 1, 2$, $j = 1, 2$. From this, two parameters $w'_{\tilde{B}_i}$ and $\delta_{\tilde{B}}$ are computed using Equation (3.16), and are then utilised to construct $\tilde{B}'$ from Equation (3.15), resulting in the intermediate rule $\tilde{A}'_1 \wedge \tilde{A}'_2 \Rightarrow \tilde{B}'$.

4.2.4 Making Scale, Move and Height Transformations

The scale rates s^V_{jp} ($j = 1, 2$, $p = 0, 1$) for scaling the support and nucleus of $\tilde{A}'^V_j$ with respect to $\tilde{A}^{*V}_j$ are calculated by

$$s^V_{j0} = \frac{\tilde{a}^{*V}_{j3} - \tilde{a}^{*V}_{j0}}{\tilde{a}'^V_{j3} - \tilde{a}'^V_{j0}}, \quad s^V_{j1} = \frac{\tilde{a}^{*V}_{j2} - \tilde{a}^{*V}_{j1}}{\tilde{a}'^V_{j2} - \tilde{a}'^V_{j1}} \tag{4.7}$$

resulting in $\tilde{A}''^V_j$, $V = L, U$. The scale ratios $\mathbb{S}^V_{jq}$ ($j = 1, 2$, $q = 1$), which represent the actual increase of the ratio between the support and the nucleus, are then utilised to further modify $\tilde{A}''^V_j$ to avoid the nucleus of the resultant set becoming wider than the support by

$$\mathbb{S}_{j1}^{V} = \begin{cases} \dfrac{\frac{\tilde{a}_{j2}^{*V}-\tilde{a}_{j1}^{*V}}{\tilde{a}_{j3}^{*V}-\tilde{a}_{j0}^{*V}} - \frac{\tilde{a}_{j2}^{'V}-\tilde{a}_{j1}^{'V}}{\tilde{a}_{j3}^{'V}-\tilde{a}_{j0}^{'V}}}{1-\frac{\tilde{a}_{j2}^{'V}-\tilde{a}_{j1}^{'V}}{\tilde{a}_{j3}^{'V}-\tilde{a}_{j0}^{'V}}} & \text{if } s_{j1}^{V} \geq s_{j0}^{V} \\ \dfrac{\frac{\tilde{a}_{j2}^{*V}-\tilde{a}_{j1}^{*V}}{\tilde{a}_{j3}^{*V}-\tilde{a}_{j0}^{*V}} - \frac{\tilde{a}_{j2}^{'V}-\tilde{a}_{j1}^{'V}}{\tilde{a}_{j3}^{'V}-\tilde{a}_{j0}^{'V}}}{\frac{\tilde{a}_{j2}^{'V}-\tilde{a}_{j1}^{'V}}{\tilde{a}_{j3}^{'V}-\tilde{a}_{j0}^{'V}}} & \text{if } s_{j0}^{V} > s_{j1}^{V} \end{cases} \tag{4.8}$$

where $j = 1, 2$, $V = L, U$. Scale transformation is then applied to generate $\tilde{B}^{''V}$ from $\tilde{B}^{'V}$ using $s_{j0}^{'V}$ and $s_{j1}^{'V}$ under the conditions $\mathbb{S}_{j1}^{'V} = \mathbb{S}_{j1}^{V}$ and $s_{j0}^{'V} = s_{j0}^{V}$. The scale rates $s_{j1}^{'V}$ of the nucleus of $\tilde{B}^{'V}$ are calculated by

$$s_{j1}^{'V} = \begin{cases} s_{j0}^{'V} * (\mathbb{S}_{j1}^{V} * \dfrac{\tilde{b}_{3}^{'V} - \tilde{b}_{0}^{'V}}{\tilde{b}_{2}^{'V} - \tilde{b}_{1}^{'V}} - \mathbb{S}_{j1}^{V} + 1) & \text{if } s_{j1}^{V} \geq s_{j0}^{V} \\ s_{j1}^{V} & \text{if } s_{j0}^{V} > s_{j1}^{V} \end{cases} \tag{4.9}$$

where $j = 1, 2$, $V = L, U$.

The similarity degree is then measured by the use of the move ratios $\mathbb{M}_{jr}^{V}$ ($j = 1, 2$, $r = 0$). By the use of $\mathbb{M}_{j0}^{V}$, $\tilde{A}_{j}^{''V}$ is moved so that the transformed set exactly matches the shape of $\tilde{A}_{j}^{*V}$. Since $r = 0$, there are no sub-moves in the transformation process, Equation (3.20) is thus not able to be used for calculation. Instead, the move ratios $\mathbb{M}_{j0}^{V}$ ($j = 1, 2$) are computed as follows:

$$\mathbb{M}_{j0}^{V} = \begin{cases} \dfrac{\tilde{a}_{j0}^{*V} - \tilde{a}_{j0}^{''V}}{\frac{\tilde{a}_{j1}^{''V}-\tilde{a}_{j0}^{''V}}{3}} & \text{if } \tilde{a}_{j0}^{*V} \geq \tilde{a}_{j0}^{''V} \\ \dfrac{\tilde{a}_{j0}^{*V} - \tilde{a}_{j0}^{''V}}{\frac{\tilde{a}_{j3}^{''V}-\tilde{a}_{j2}^{''V}}{3}} & \text{if } \tilde{a}_{j0}^{''V} > \tilde{a}_{j0}^{*V} \end{cases} \tag{4.10}$$

where $j = 1, 2$, $V = L, U$.

The similarity degree is further reinforced using the height rates h_{jo} ($j = 1, 2$, $o = 1, 2$), which are calculated by

$$h_{j1} = \frac{\tilde{H}_{\tilde{A}_{j1}}^{*L}}{\tilde{H}_{\tilde{A}_{j1}}^{'L}}, \quad h_{j2} = \frac{\tilde{H}_{\tilde{A}_{j2}}^{*L}}{\tilde{H}_{\tilde{A}_{j2}}^{'L}} \tag{4.11}$$

where $j = 1, 2$.

4.2.5 Deriving Interpolated Conclusion

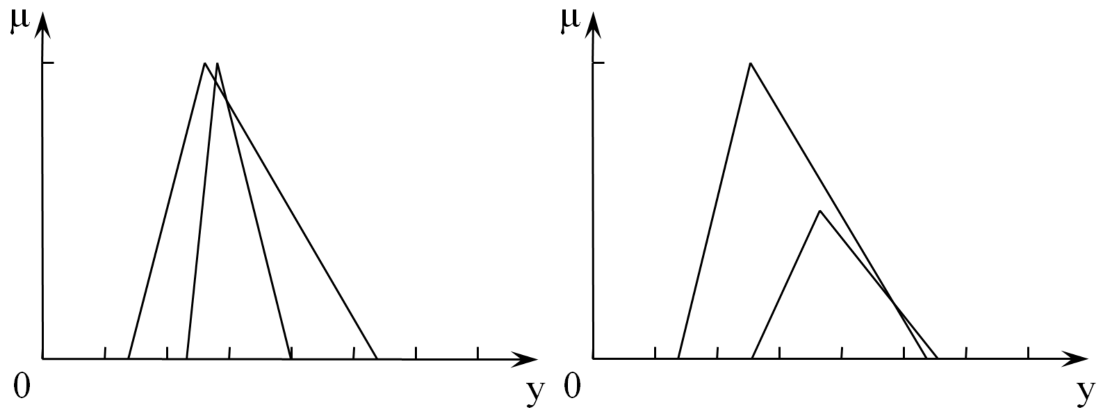
Finally, the interpolated conclusion $\tilde{B}^*$ is estimated using the aggregated parameters $s^V_{\tilde{B}_0}$, $s^V_{\tilde{B}_1}$, $\mathbb{S}^V_{\tilde{B}_1}$, $\mathbb{M}^V_{\tilde{B}_0}$, $h_{\tilde{B}_1}$ and $h_{\tilde{B}_2}$, $V = L, U$, where

$$s^V_{\tilde{B}_0} = \frac{1}{2}(s'^V_{10} + s'^V_{20}), \;\; s^V_{\tilde{B}_1} = \frac{1}{2}(s'^V_{11} + s'^V_{21}), \;\; \mathbb{S}^V_{\tilde{B}_1} = \frac{1}{2}(\mathbb{S}'^V_{11} + \mathbb{S}'^V_{21}) \tag{4.12}$$

$$\mathbb{M}^V_{\tilde{B}_0} = \frac{1}{2}(\mathbb{M}^V_{10} + \mathbb{M}^V_{20}), \;\; h_{\tilde{B}_1} = \frac{1}{2}(h_{11} + h_{21}), \;\; h_{\tilde{B}_2} = \frac{1}{2}(h_{12} + h_{22}) \tag{4.13}$$

4.3 Modified Procedure

A key concept in T-FRI is the Rep of a given conventional FS. In RF rule interpolation, the overall location of an RF set is not just based on both the location of the LA and that of the UA. Intuitively, the shape of the LA should not violate the shape of the UA. Unfortunately, this is not guaranteed in cases where the normal points or base points are identical, where non-intuitive results may be interpolated using the above algorithm. As illustrated in Figure 4.3, the LA of the interpolated result is sometimes greater than its UA. Having recognised this, in order to obtain intuitive interpolated conclusions for RF sets when implementing the framework, the relative location between the LA and UA should also be taken into consideration. Therefore, a modified procedure is introduced below.



(a) Interpolation with identical normal points (b) Interpolation with identical right base points

Figure 4.3: Two interpolated results

For simplicity, only rules involving trapezoidal-shaped RF sets with single antecedents are discussed here (though the work is applicable to multi-antecedent rules as to be shown later). Suppose that an intermediate rule $\tilde{A}' \Rightarrow \tilde{B}'$ and the similarity degree between $\tilde{A}'$ and the observation $\tilde{A}^*$ have already been calculated, resulting in the scaled intermediate term $\tilde{A}''$. The scaled intermediate term $\tilde{B}''$ is then obtained from $\tilde{B}'$ by the use of the parameters $s_0^{'V}$, $s_1^{'V}$ and $\mathbb{S}_1^{'V}$ in the existing algorithm. The following procedure is used to modify $\tilde{B}''$ into $\tilde{B}_c''$ by an aggregation operation. This offers combined information about the temporary values $\tilde{B}^{''L}$ and $\tilde{B}^{''U}$ that are obtained by two reversed processes, using $s_0^{'V}$, $s_1^{'V}$ and $\mathbb{S}_1^{'V}$, $V \in \{L, U\}$.

4.3.1 Scale Transformation

These scale rates are employed to transform $\tilde{B}^{'L}$ and $\tilde{B}^{'U}$ into $\tilde{B}^{''L}$ and $\tilde{B}^{''U}$, respectively. The relative location between $\tilde{B}^{'L}$ and $\tilde{B}^{'U}$ of $\tilde{B}'$ is defined by the *relative location factor* θ:

$$\theta = \frac{\tilde{B}^{'L}}{\tilde{B}^{'U}} \tag{4.14}$$

which is computed by

$$\theta_k = \frac{\tilde{b}_k^{'L}}{\tilde{b}_k^{'U}}, \quad k = 0, \cdots, 3 \tag{4.15}$$

Note that the relative location between $\tilde{B}^{''L}$ and $\tilde{B}^{''U}$ should be associated with that between $\tilde{B}^{'L}$ and $\tilde{B}^{'U}$. Otherwise, it will result in the non-intuitive interpolated conclusions, i.e., $\tilde{B}^{''L}/\tilde{B}^{''U} \neq \theta$. The relative location factor θ is thus used to maintain the relative location both before and after the scale transformation.

4.3.2 Modification

The modification process is applied such that:

$$\theta = \frac{\tilde{B}^{''L}}{\tilde{B}_n^{''U}} = \frac{\tilde{B}_n^{''L}}{\tilde{B}^{''U}} \tag{4.16}$$

Given $\tilde{B}^{''L}$, $\tilde{B}^{''U}$ can be modified into a "new" $\tilde{B}_n^{''U}$ using θ (where the subscript n stands for "new"). Similarly, given $\tilde{B}^{''U}$, $\tilde{B}^{''L}$ can be modified into a "new" $\tilde{B}_n^{''L}$ by the same θ. These two reversed processes are described as follows:

$$\begin{cases} \tilde{B}^{'L} \xrightarrow{s_0^{'L},\, s_1^{'L},\, \mathbb{S}_1^{'L}} \tilde{B}^{''L} \xrightarrow{\theta} \tilde{B}_n^{''U} \\ \tilde{B}^{'U} \xrightarrow{s_0^{'U},\, s_1^{'U},\, \mathbb{S}_1^{'U}} \tilde{B}^{''U} \xrightarrow{\theta} \tilde{B}_n^{''L} \end{cases} \tag{4.17}$$

which result in two pairs of fuzzy terms. An aggregation is then used to obtain the combined scaled intermediate term $\tilde{B}_c''$.

4.3.3 Aggregation

The combined $\tilde{B}_c''^L$ and $\tilde{B}_c''^U$ of $\tilde{B}_c''$ (where the subscript c stands for "combined") are then computed as the average of the corresponding two terms, respectively, such that

$$\tilde{B}_c''^V = \frac{\tilde{B}''^V + \tilde{B}_n''^V}{2}, \quad V = L, U \tag{4.18}$$

Note that the aggregation is herein defined this way because both terms are equally important. However, alternative definitions may be introduced for this, but they may complicate the calculation involved. The relative location between $\tilde{B}_c''^L$ and $\tilde{B}_c''^U$ remains the same θ, i.e., $\tilde{B}_c''^L / \tilde{B}_c''^U = \theta$. Therefore, the intuitive interpolated conclusions can be ensured with the use of the relative location factor θ.

Proof. With Equations (4.16) and (4.18),

$$\begin{aligned}\frac{\tilde{B}_c''^L}{\tilde{B}_c''^U} &= \frac{\frac{1}{2}(\tilde{B}''^L + \tilde{B}_n''^L)}{\frac{1}{2}(\tilde{B}''^U + \tilde{B}_n''^U)} \\ &= \frac{\frac{1}{2}(\theta\tilde{B}_n''^U + \theta\tilde{B}''^U)}{\frac{1}{2}(\tilde{B}''^U + \tilde{B}_n''^U)} \\ &= \theta\end{aligned}$$

□

Similarly, the final interpolated conclusion can also be modified from $\tilde{B}^*$ to $\tilde{B}_c^*$ using the same θ to maintain the relative location both before and after the move transformation. In order to avoid duplication, the mathematical details of the modification of the move transformation are omitted here. An example that considers an extreme case is given here to show the improvement as compared to the non-intuitive result of Figure 4.3a.

Example 4.1. The observation and the two closest rules associated with a single antecedent variable are described as follows, where each triangular RF set under consideration has identical normal points. The involved RF sets are listed in Table 4.1.

The antecedent and consequent of the intermediate rule $\tilde{A}' =< (6.74, 8.46, 8.46, 8.96; 1, 1), (5.74, 8.46, 8.46, 9.70; 1, 1) >$ and $\tilde{B}' =< (6.46, 6.96, 6.96, 7.70; 1, 1),$

Table 4.1: Involved RF sets for Example 4.1

Attribute values	$\tilde{A}_1 =< (1, 3.5, 3.5, 4; 1, 1), (0, 3.5, 3.5, 5; 1, 1) >$
	$\tilde{A}_2 =< (12, 13, 13, 13.5; 1, 1), (11, 13, 13, 14; 1, 1) >$
	$\tilde{B}_1 =< (1.5, 2, 2, 3; 1, 1), (0, 2, 2, 5; 1, 1) >$
	$\tilde{B}_2 =< (11, 11.5, 11.5, 12; 1, 1), (10, 11.5, 11.5, 13; 1, 1) >$
Observation	$\tilde{A}^* =< (6.5, 8, 8, 9.5; 1, 1), (6, 8, 8, 10; 1, 1) >$

$(5.22, 6.96, 6.96, 9.18; 1, 1) >$ are calculated first. For scale transformation, the support scale rates $s_0^{'L} = 1.35$ and $s_0^{'U} = 1.01$ are computed in the first instance by transforming $\tilde{A}'$ to $\tilde{A}''$, and are then used to construct $\tilde{B}''$. The parameters $\theta_0 = 1.24$, $\theta_1 = 1$, $\theta_2 = 1$ and $\theta_3 = 0.84$ are calculated using Equation (4.15), resulting in the newly modified $\tilde{B}''_n$. The combined $\tilde{B}''_c =< (6.34, 6.95, 6.95, 7.83; 1, 1), (5.13, 6.95, 6.95, 9.33; 1, 1) >$ is then obtained from the average of $\tilde{B}''$ and $\tilde{B}''_n$, using Equations (4.16) and (4.18). Similarly, the same θ_0, θ_1, θ_2 and θ_3 are used to modify $\tilde{B}^*$, which is constructed from $\tilde{B}''_c$ using the move transformation, resulting in the final interpolated conclusion $\tilde{B}^*_c =< (6.49, 6.70, 6.70, 7.93; 1, 1), (5.24, 6.70, 6.70, 9.45; 1, 1) >$, as shown in Figure 4.4.

It can be seen that the normal points of the interpolated result are identical also. In contrast, it is not the case in Figure 4.3a. This shows that the modified procedure is an effective improvement for avoiding non-intuitive interpolated conclusions.

In order to further explain the computation involved, a general example which concerns an interpolation using multiple rules with multiple antecedent variables is also provided below.

Example 4.2. Suppose that four rules each involving three antecedents have been chosen as the closest rules to determine the interpolated result. All conditions are shown in Table 4.2. For the first antecedent, the distances between $\tilde{A}_{k1}$, $k = 1, 2, 3, 4$ and the observed $\tilde{A}^*_1$ are calculated by Equation (3.9). The weights are respectively calculated and normalised using Equations (3.11) and (3.12), resulting in the new weights of 0.38, 0.17, 0.11 and 0.34. The normalised weights together with the parameter $\delta_{\tilde{A}_1} = -0.14$, which is computed by Equation (3.14), are then used to generate the required intermediate FS $\tilde{A}'_1 =< (5.55, 6.36, 7.27, 7.91; 0.7, 0.7), (4.63, 5.91, 7.74, 9.10; 1, 1) >$ according to Equation (3.13). $\tilde{A}'_2$ and $\tilde{A}'_3$ can then be calculated in the same way. For the consequence, the average weights of 0.37, 0.21, 0.21

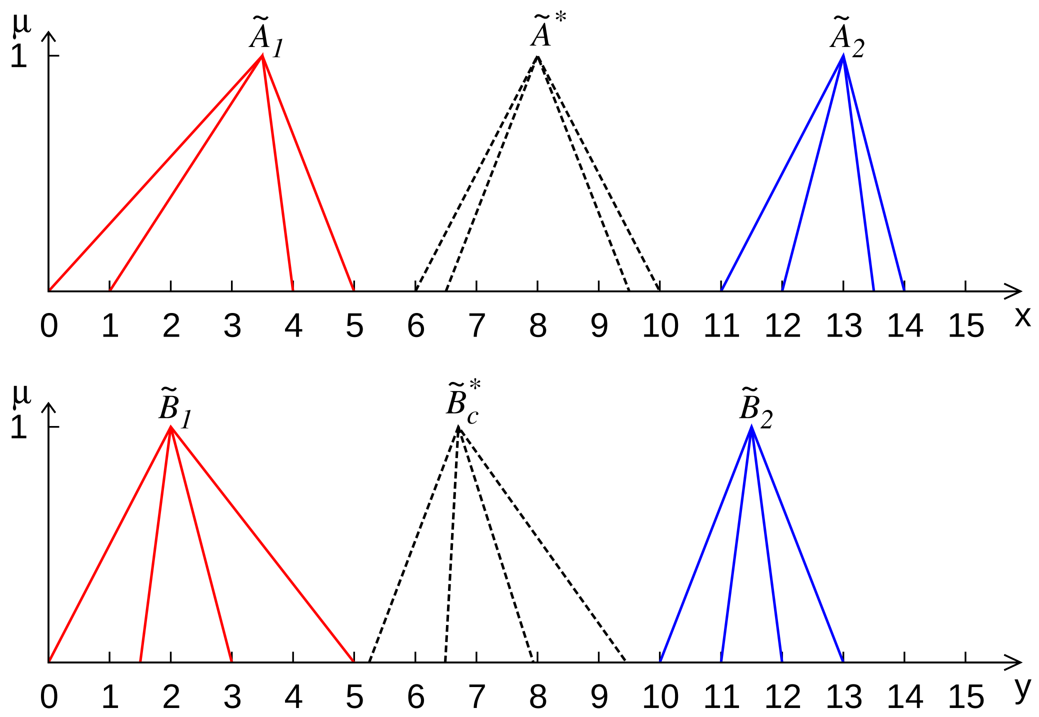


Figure 4.4: Example for triangular RF sets with identical normal points

and 0.21, and the average parameter $\delta_{\tilde{B}} = -0.15$ can be calculated using Equation (3.16). From this, the intermediate output $\tilde{B}' =< (6.61, 7.71, 8.80, 9.40; 0.6, 0.6), (6.11, 7.61, 9.01, 10.11; 1, 1) >$ is obtained with respect to Equation (3.15). The average of three support scale rates (0.63, 1.01 and 0.83) of the LAs and the average of three nucleus scale rates (0.55, 0.47 and 0.55) of the LAs are computed according to Equations (4.7), (4.8) and (4.9), resulting in $s^L_{\tilde{B}_0} = 0.82$ and $s^L_{\tilde{B}_1} = 0.52$, forming the aggregated scale rates of the LA. The aggregated scale rates of the UA $s^U_{\tilde{B}_0} = 0.92$ and $s^U_{\tilde{B}_1} = 0.33$ can then be generated following the same procedure. Similarly, the aggregated move ratios $\mathbb{M}^L_{\tilde{B}_0} = 0.49$ and $\mathbb{M}^U_{\tilde{B}_0} = -0.04$ are calculated from three move ratios (0.06, 0.91 and 0.50) of the LAs and three move ratios (0.27, −0.37 and −0.02) of the UAs using Equation (4.10). These, together with the aggregated height rates, namely, the average $h_{\tilde{B}_1} = 1$ and $h_{\tilde{B}_2} = 1$ of the two pairs of height rates (1, 1 and 1) and (1, 1 and 1) from Equation (4.11), are employed to transform $\tilde{B}'$ to achieve the final result $\tilde{B}^*_c =< (6.90, 7.90, 8.26, 9.30; 0.6, 0.6), (6.38, 7.80, 8.46, 10.00; 1, 1) >$, with $\theta_0 = 1.08$, $\theta_1 = 1.01$, $\theta_2 = 0.98$ and $\theta_3 = 0.93$. The interpolated result is illustrated in Figure 4.5. The resultant interpolated RF value has a clear intuitive appeal.

Table 4.2: Four closest rules for observation

	Antecedents	Consequence
Rule 1	$\tilde{A}_{11} =< (1, 1.5, 2.5, 3; 0.7, 0.7), (0, 1, 3, 4.5; 1, 1) >$	
	$\tilde{A}_{12} =< (2, 2.5, 3.5, 4; 0.5, 0.5), (1, 2, 3.5, 5; 1, 1) >$	$\tilde{B}_1 =< (0.5, 1.5, 3, 3.5; 0.6, 0.6), (0, 1.5, 3, 4; 1, 1) >$
	$\tilde{A}_{13} =< (1, 2, 3, 4; 0.6, 0.6), (0.5, 1, 4, 5; 1, 1) >$	
Rule 2	$\tilde{A}_{21} =< (15.5, 16.5, 18, 19; 0.7, 0.7), (15, 16, 18, 20; 1, 1) >$	
	$\tilde{A}_{22} =< (11, 11.5, 13, 14; 0.5, 0.5), (10, 11, 13, 15; 1, 1) >$	$\tilde{B}_2 =< (21, 22, 22.5, 23; 0.6, 0.6), (20.5, 22, 23, 24; 1, 1) >$
	$\tilde{A}_{23} =< (20.5, 21, 22, 23; 0.6, 0.6), (20, 21, 22, 23.5; 1, 1) >$	
Rule 3	$\tilde{A}_{31} =< (21, 22, 23, 24; 0.7, 0.7), (20, 22, 24, 25; 1, 1) >$	
	$\tilde{A}_{32} =< (16, 17, 18, 19; 0.5, 0.5), (15, 17, 19, 20.5; 1, 1) >$	$\tilde{B}_3 =< (15, 17, 18, 19; 0.6, 0.6), (14.5, 17, 18, 20; 1, 1) >$
	$\tilde{A}_{33} =< (12, 12.5, 13.5, 14; 0.6, 0.6), (11, 11.5, 13.5, 15; 1, 1) >$	
Rule 4	$\tilde{A}_{41} =< (11, 12, 12.5, 13; 0.7, 0.7), (10, 11.5, 13, 14; 1, 1) >$	
	$\tilde{A}_{42} =< (21.5, 22, 23, 23.5; 0.5, 0.5), (21, 22, 23, 24; 1, 1) >$	$\tilde{B}_4 =< (12, 12.5, 13.5, 14; 0.6, 0.6), (11.5, 12, 14, 14.5; 1, 1) >$
	$\tilde{A}_{43} =< (17, 18, 18.5, 19; 0.6, 0.6), (16, 17, 19, 20; 1, 1) >$	
Observation	$\tilde{A}_1^* =< (6, 6.5, 7, 7.5, ; 0.7, 0.7), (5, 6, 7, 9; 1, 1) >$	
	$\tilde{A}_2^* =< (6, 6.5, 8, 8.5; 0.5, 0.5), (5, 6, 8, 9; 1, 1) >$	
	$\tilde{A}_3^* =< (7, 7.5, 8, 9; 0.6, 0.6), (6, 7, 8, 10; 1, 1) >$	

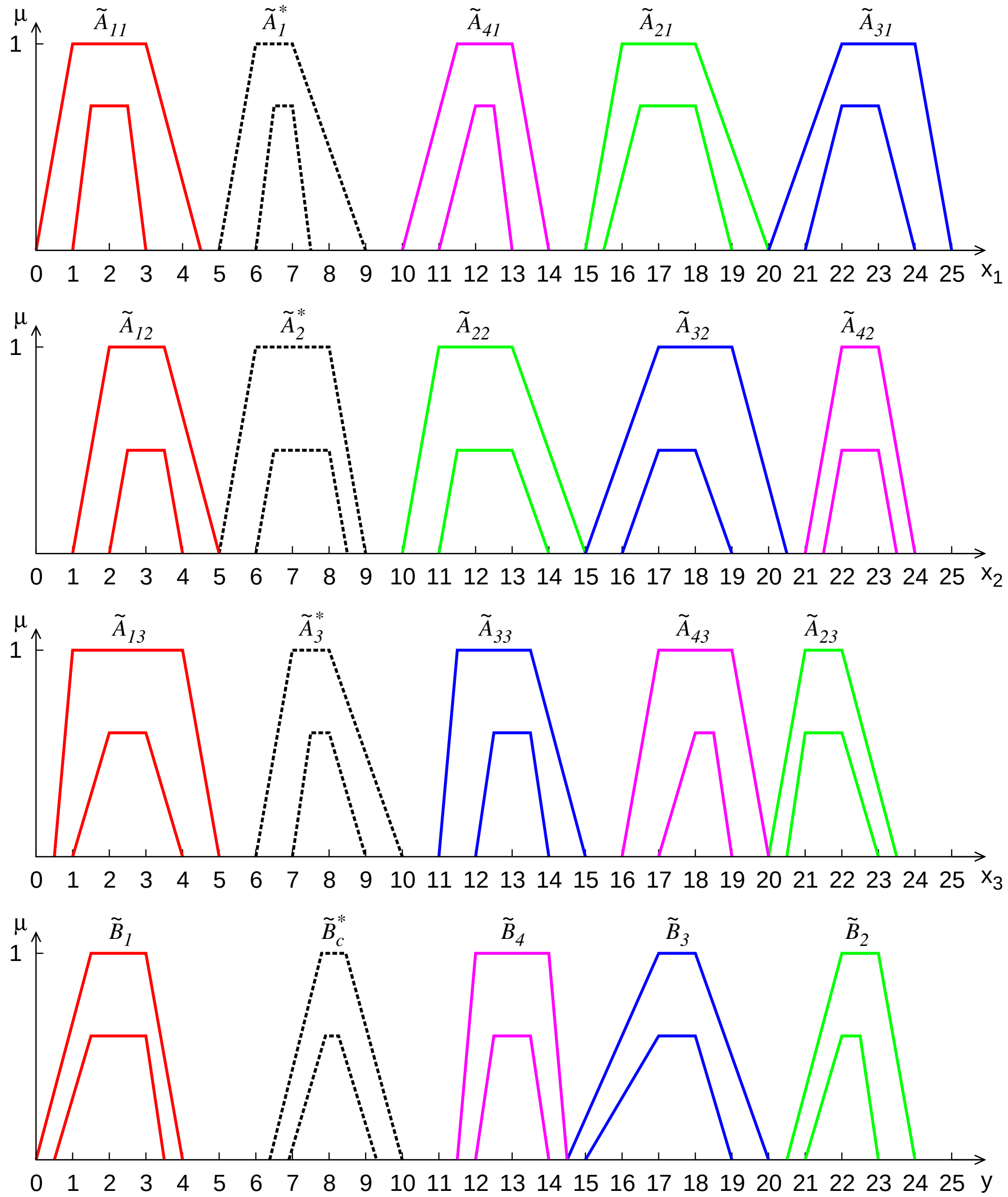


Figure 4.5: Example for RF interpolation with multiple rules

4.4 Extrapolation

The extension of the above to perform extrapolation is readily attainable. If all the chosen closest rules lie on one side of the observation, the interpolation problem becomes extrapolation. Both choosing the closest rules and constructing the intermediate rule are carried out in the exactly same way as those procedures for interpolation as described in Section 4.2.

An example follows to explain the process of RF rule extrapolation, which in essence is the same as that for RF rule interpolation. Suppose that only the second, third and fourth rules in Example 4.2 are considered, given the case the interpolation process becomes extrapolation using three rules.

Example 4.3. Three rules $\tilde{A}_{k1} \wedge \tilde{A}_{k2} \wedge \tilde{A}_{k3} \Rightarrow \tilde{B}_k$, $k = 2, 3, 4$ and the observations $\tilde{A}^*_1$, $\tilde{A}^*_2$, $\tilde{A}^*_3$ as given in Table 4.2 are used to carry out fuzzy extrapolation in this example. For the first attribute $\tilde{A}_1$, the normalised weights of $\tilde{A}_{k1}$, $k = 2, 3, 4$ are computed to become 0.27, 0.18 and 0.55. They are used to obtain an RF term $\tilde{A}^{IFT}_1$ by Equation (3.13). According to Equation (3.14), $\delta_{\tilde{A}_1} = -0.45$ is obtained. $\tilde{A}^{IFT}_1$ and $\delta_{\tilde{A}_1}$ are then utilised to generate the intermediate RF set $\tilde{A}'_1 =< (5.45, 6.45, 7.31, 8.04; 0.7, 0.7), (4.59, 6.04, 7.77, 9.04; 1, 1) >$. Similarly, $\tilde{A}'_2$ and $\tilde{A}'_3$ can be constructed following the above procedure. For the consequence, the intermediate output $\tilde{B}' =< (7.50, 8.66, 9.49, 10.16; 0.6, 0.6), (7.00, 8.50, 9.83, 10.99; 1, 1) >$ is computed using the average weights of 0.34, 0.33 and 0.33, and the average $\delta_{\tilde{B}}$ of -0.47, with respect to Equations (3.15) and (3.16). The average $s^L_{\tilde{B}_0} = 0.81$, $s^L_{\tilde{B}_1} = 0.42$, $s^U_{\tilde{B}_0} = 0.92$ and $s^U_{\tilde{B}_1} = 0.40$ are calculated in terms of Equations (4.7), (4.8) and (4.9). According to Equations (4.10) and (4.11), the average $\mathbb{M}^L_{\tilde{B}_0} = 0.43$, $\mathbb{M}^U_{\tilde{B}_0} = 0.003$, $h_{\tilde{B}_1} = 1$ and $h_{\tilde{B}_2} = 1$ can then be obtained as well. The final result $\tilde{B}^*_c =< (7.80, 8.81, 9.00, 10.05; 0.6, 0.6), (7.28, 8.64, 9.32, 10.87; 1, 1) >$ is therefore constructed with $\theta_0 = 1.07$, $\theta_1 = 1.02$, $\theta_2 = 0.97$ and $\theta_3 = 0.92$, which are computed by Equation (4.15), as shown in Figure 4.6. Again, the result has an intuitive appeal.

Note that the result in Figure 4.6 is closer to the right three rules than the result in Figure 4.5. This is because the left rule is not considered in Example 4.3, resulting in higher weights being assigned to the three rules on the right side of the observation.

4.5 Compatibility with T-FRI

As the concept of RF sets extends from conventional FSs, the RF interpolation extends from the existing T-FRI. When the involved higher order uncertainty disappears, namely, the LA coincides with the LU, an RF set degenerates to a conventional FS. If all the considered sets in the implementation of interpolation/extrapolation are conventional FSs, the results obtained by the proposed approach should therefore be

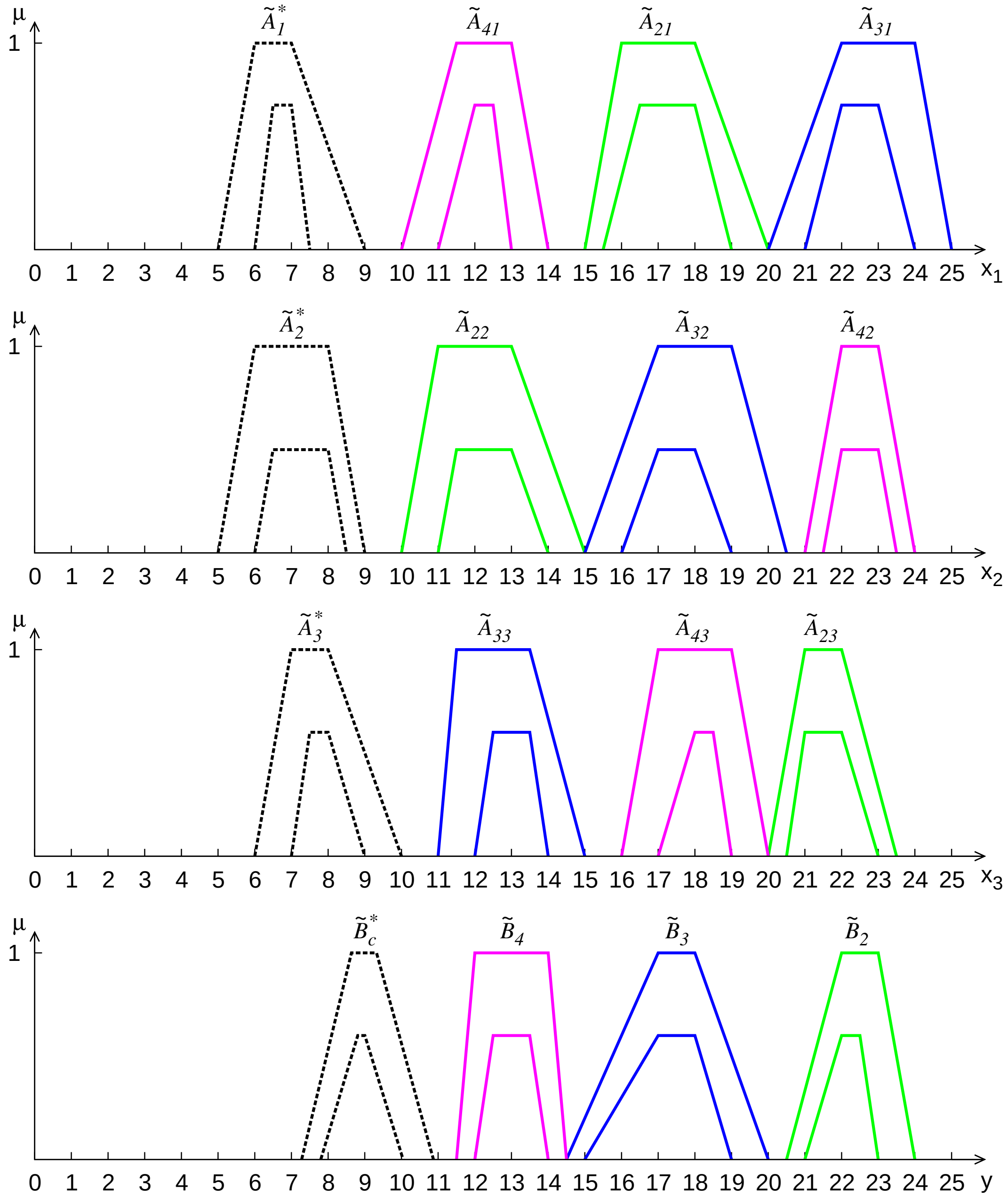


Figure 4.6: Example for RF extrapolation with multiple rules

identical to those by T-FRI. For this reason, theorems are provided in this section in order to verify that the RF approach is indeed compatible with the original one.

The differences between the RF approach and the T-FRI are reflected from three aspects: the calculation of Rep, the height transformation and the modified procedure. In order to ensure that identical conclusion is retained when all RF sets degenerate to conventional (type-1) FSs, the key point is to ensure that an identical intermediate rule is obtained accordingly. Similarly, in order to ensure the identical intermediate rule, the key point is to have an identical distance measure result between two types of representation. For simplicity, only trapezoidal sets are considered here.

4.5.1 Initial Condition

Assume a trapezoidal RF set $\tilde{A}$ is represented as $\tilde{A} =< \tilde{A}^L, \tilde{A}^U >$, where $\tilde{A}^L = (\tilde{a}_0^L, \tilde{a}_1^L, \tilde{a}_2^L, \tilde{a}_3^L; \tilde{H}^L_{\tilde{A}_1}, \tilde{H}^L_{\tilde{A}_2})$, $\tilde{A}^U = (\tilde{a}_0^U, \tilde{a}_1^U, \tilde{a}_2^U, \tilde{a}_3^U; \tilde{H}^U_{\tilde{A}_1}, \tilde{H}^U_{\tilde{A}_2})$. $\tilde{A}^L$ coincides with $\tilde{A}^U$, namely, $\tilde{a}_i^L = \tilde{a}_i^U$ $(i = 0, \cdots, 3)$ and $\tilde{H}^L_{\tilde{A}_E} = \tilde{H}^U_{\tilde{A}_E}$ $(E = 1, 2)$. That is, all the considered sets in this subsection are conventional FSs. The conclusion derived from T-FRI is denoted by B^*_T, and this subscript applies to other intermediate results as well. The purpose of the comparison is to show that the same conclusions will be obtained from these two approaches, i.e., if $\tilde{A}_{ij} = A_{ij}$, $\tilde{A}^*_j = A^*_j$ and $\tilde{B}_i = B_i$, then $\tilde{B}^* = B^*_T$, $i \in \{1, \cdots, N\}$, $j \in \{1, \cdots, M\}$.

Theorem 4.1. For two HOFSs $\tilde{A}_1$ and $\tilde{A}_2$, if $\tilde{A}^L_1 = \tilde{A}^U_1$ and $\tilde{A}^L_2 = \tilde{A}^U_2$, then the distance measure $d(\tilde{A}_1, \tilde{A}_2) = d(A_1, A_2)_T$.

Proof. The weight is herein determined by the arithmetic average operator. With Equation (3.5),

$$\mathrm{Rep}(\tilde{A}^L_1)_x = \mathrm{Rep}(\tilde{A}^U_1)_x = \mathrm{Rep}(A_1)_T, \quad \mathrm{Rep}(\tilde{A}^L_1)_y = \mathrm{Rep}(\tilde{A}^U_1)_y = 1/4$$

With Equations (3.6) and (3.7),

$$f^L_{A_1} = f^U_{A_1}, \quad w^L_{A_1} = w^U_{A_1} = 1/2$$

With Equation (3.8),

$$\mathrm{Rep}(\tilde{A}_1) = \mathrm{Rep}(\tilde{A}^L_1)_x + 1/4$$

The calculation for A_2 follows the same procedure. In this case, with Equation (3.9),

$$\begin{aligned} d(\tilde{A}_1, \tilde{A}_2) &= d(\mathrm{Rep}(\tilde{A}_1), \mathrm{Rep}(\tilde{A}_2)) \\ &= \mathrm{Rep}(\tilde{A}^L_1)_x + 1/4 - \mathrm{Rep}(\tilde{A}^L_2)_x - 1/4 \\ &= \mathrm{Rep}(A_1)_T - \mathrm{Rep}(A_2)_T \\ &= d(A_1, A_2)_T \end{aligned}$$

□

4.5.2 Intermediate Rule

The implementation of the intermediate rule can then be verified using Theorem 4.1.

Theorem 4.2. Given the observation A_j^*, if R_i is a rule involving M antecedent variables in the rule base, then the distance $d_i = (d_i)_T$.

Proof. With Equation (3.9) and Theorem 4.1,

$$\begin{aligned} d_{ij} &= d(\tilde{A}_{ij}, \tilde{A}_j^*) \\ &= d(A_{ij}, A_j^*)_T = (d_{ij})_T \end{aligned}$$

where $j \in \{1, \cdots, M\}$. With Equation (3.10),

$$\begin{aligned} d_i &= \sqrt{\sum_{j=1}^{M} (\frac{d_{ij}}{\max_j - \min_j})^2} \\ &= \sqrt{\sum_{j=1}^{M} (\frac{(d_{ij})_T}{\max_j - \min_j})^2} = (d_i)_T \end{aligned}$$

□

The above proof shows that the identical N ($N \geq 2$) rules which have minimal distances will be chosen as the closest N rules from these two approaches.

Theorem 4.3. Given $\tilde{A}_1^* \wedge \cdots \tilde{A}_j^* \wedge \cdots \tilde{A}_M^*$, if $\tilde{A}_{i1} \wedge \cdots \tilde{A}_{ij} \wedge \cdots \tilde{A}_{iM} \Rightarrow \tilde{B}_i$, $i \in \{1, \cdots, N\}$, $j \in \{1, \cdots, M\}$, are N chosen closest rules, then the intermediate rule $\tilde{A}'_1 \wedge \cdots \tilde{A}'_j \wedge \cdots \tilde{A}'_M \Rightarrow \tilde{B}'$ is identical to that of T-FRI.

Proof. With Equations (3.13), (3.14) and (3.12),

$$\begin{aligned} w'_{\tilde{A}_{ij}} &= \frac{\frac{1}{d_{ij}}}{\sum_{i=1}^{N} \frac{1}{d_{ij}}} = \frac{\frac{1}{(d_{ij})_T}}{\sum_{i=1}^{N} \frac{1}{(d_{ij})_T}} = (w'_{A_{ij}})_T, \\ \tilde{A}_j^{IFT} &= \sum_{i=1}^{N} w'_{\tilde{A}_{ij}} \tilde{A}_{ij} = \sum_{i=1}^{N} (w'_{A_{ij}})_T A_{ij} = (A_j^{IFT})_T, \\ \delta_{\tilde{A}_j} &= \frac{\mathrm{Rep}(\tilde{A}_j^*) - \mathrm{Rep}(\tilde{A}_j^{IFT})}{\max_j - \min_j} = \frac{\mathrm{Rep}(A_j^*) - \mathrm{Rep}(A_j^{IFT})_T}{\max_j - \min_j} = (\delta_{A_j})_T, \\ \therefore \tilde{A}'_j &= \tilde{A}_j^{IFT} + \delta_{\tilde{A}_j}(\max_j - \min_j) \\ &= (A_j^{IFT})_T + (\delta_{A_j})_T(\max_j - \min_j) = (A'_j)_T \end{aligned}$$

Similarly, with Equations (3.15) and (3.16),

$$w'_{\tilde{B}_i} = \frac{1}{M}\sum_{j=1}^{M} w'_{\tilde{A}_{ij}} = \frac{1}{M}\sum_{j=1}^{M}(w'_{A_{ij}})_T = (w'_{B_i})_T,$$

$$\tilde{B}^{IFT} = \sum_{i=1}^{N} w'_{\tilde{B}_i}\tilde{B}_i = \sum_{i=1}^{N}(w'_{B_i})_T B_i = B_T^{IFT},$$

$$\delta_{\tilde{B}} = \frac{1}{M}\sum_{j=1}^{M}\delta_{\tilde{A}_j} = \frac{1}{M}\sum_{j=1}^{M}(\delta_{A_j})_T = (\delta_B)_T,$$

$$\begin{aligned}\therefore \tilde{B}' &= \tilde{B}^{IFT} + \delta_{\tilde{B}}(\max - \min)\\ &= B_T^{IFT} + (\delta_B)_T(\max - \min) = B'_T\end{aligned}$$

□

4.5.3 Height Transformation

When RF sets degenerate to conventional FSs, the heights of the LA are 1 owing to the normality. The purpose of the height transformation is now to keep the heights of the resulting consequence the same. Note that for the sake of simplicity, only fuzzy terms involving one single antecedent variable are considered in Theorem 4.4 (and also in Theorem 4.5 to be presented in the next subsection). However, the underlying ideas can be easily used to address more general cases.

Theorem 4.4. Given two adjacent rules $\tilde{A}_1 \Rightarrow \tilde{B}_1$ and $\tilde{A}_2 \Rightarrow \tilde{B}_2$, and an observation $\tilde{A}^*$, if $\tilde{H}^V_{\tilde{A}_{1E}} = \tilde{H}^V_{\tilde{A}_{2E}} = \tilde{H}^V_{\tilde{B}_{1E}} = \tilde{H}^V_{\tilde{B}_{2E}} = \tilde{H}^{*V}_{\tilde{A}_E} = 1$, $V = L, U$, $E = 1, 2$, then $\tilde{H}^*_{\tilde{B}} = (H^*_B)_T$.

Proof. With Equations (3.13) and (3.15),

$$\tilde{H}'^V_{\tilde{A}_E} = (H'^V_{A_E})_T = 1$$

With Equation (3.21),

$$\tilde{H}^{*V}_{\tilde{B}} = (H^{*V}_B)_T = 1$$

□

4.5.4 Modified Procedure

Since the algorithms for scale transformation are the same in the proposed approach and T-FRI, it can be seen that the second intermediate terms $\tilde{A}'' = A''_T$ and $\tilde{B}'' = B''_T$. The combined result $\tilde{B}''_c$ can then be verified that it is the same as B''_T by the following theorem.

Theorem 4.5. Given $\tilde{A}'' \Rightarrow \tilde{B}''$ and $A''_T \Rightarrow B''_T$, if $\tilde{A}'' = A''_T$ and $\tilde{B}'' = B''_T$, then the combined $\tilde{B}''_c = B''_T$.

Proof. With Equation (4.14),

$$\theta = \frac{\tilde{B}'^L}{\tilde{B}'^U} = 1$$

With Equation (4.16),

$$\tilde{B}''^L = \tilde{B}''^U_n = \tilde{B}''^L_n = \tilde{B}''^U$$

With Equation (4.18),

$$\begin{aligned}
\tilde{B}''^L_c &= \frac{\tilde{B}''^L + \tilde{B}''^L_n}{2} = \tilde{B}''^L \\
\tilde{B}''^U_c &= \frac{\tilde{B}''^U + \tilde{B}''^U_n}{2} = \tilde{B}''^U \\
\therefore \tilde{B}''_c &= \tilde{B}'' = B''_T
\end{aligned}$$

□

The modified procedure for move transformation can be similarly verified as well. The proof is omitted here to avoid duplication.

4.5.5 Illustrative Example

Consequently, the above theorems jointly show that the RF approach collapses to the existing T-FRI if higher order uncertainty degenerates to type-1 fuzziness. That is, the proposed RF interpolation and extrapolation extends T-FRI, addressing both the first order and higher order types of uncertainty coherently. The following example illustrates this.

Example 4.4. Consider a specific case where all of the RF sets degenerate to conventional FSs, i.e., $\tilde{A}^{*L} = \tilde{A}^{*U}$, $\tilde{A}^L_k = \tilde{A}^U_k$ and $\tilde{B}^L_k = \tilde{B}^U_k$, $k = 1, 2$. Let $\tilde{A}^*$, $\tilde{A}_1$, $\tilde{A}_2$, $\tilde{B}_1$ and $\tilde{B}_2$ be RF sets, as listed in Table 4.3.

Using the proposed approach, the interpolated conclusion $\tilde{B}^*_c =< (5.23, 5.23, 7.61, 8.32; 1, 1), (5.23, 5.23, 7.61, 8.32; 1, 1) >$ can be obtained, as shown in Figure 4.7. The details of the calculation are omitted here to avoid repetition. It follows that if all given sets are conventional FSs, then the interpolated result is the same as that achieved using the classical T-FRI.

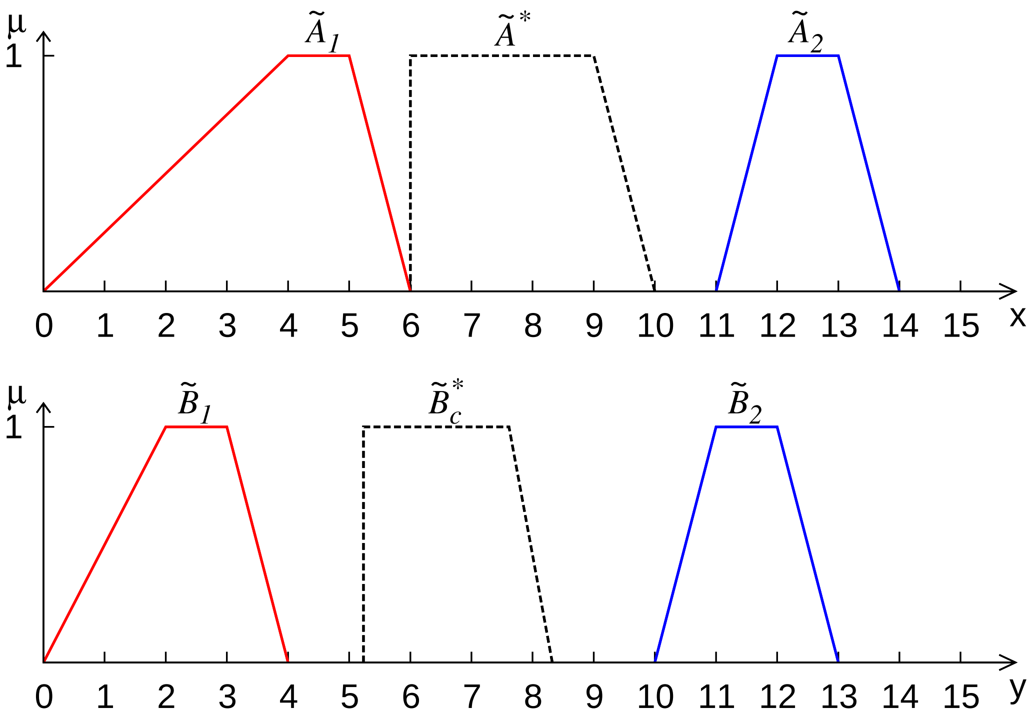


Figure 4.7: Example for conventional FSs case

Table 4.3: Involved RF sets for Example 4.4

Attribute values	$\tilde{A}_1 =< (0,4,5,6;1,1),(0,4,5,6;1,1) >$
	$\tilde{A}_2 =< (11,12,13,14;1,1),(11,12,13,14;1,1) >$
	$\tilde{B}_1 =< (0,2,3,4;1,1),(0,2,3,4;1,1) >$
	$\tilde{B}_2 =< (10,11,12,13;1,1),(10,11,12,13;1,1) >$
Observation	$\tilde{A}^* =< (6,6,9,10;1,1),(6,6,9,10;1,1) >$

4.6 Summary

This chapter has described an implementation of the proposed framework with the use of RF sets. It has introduced the concepts of lower and upper approximation MFs and presented an algorithm for RF rule interpolation, assuming that sparse rules involving RF-valued variables are available. A refinement procedure to ensure intuitive interpolated conclusions has been explained. Also, a proof has been provided to verify that the proposed approach is compatible with the original T-FRI.

Chapter 5

Implementing Framework with Type-2 Fuzzy Sets

As an extension of the conventional (type-1) FSs, type-2 FSs are finding wide applicability in rule-based fuzzy systems because of their extended power in expressing uncertainty in fuzzy modelling. Interval type-2 FSs, which are a special and simple category of type-2 FSs, are computationally simple and therefore are used in this work, to develop a type-2 FRI technique as an alternative implementation of the framework introduced in Chapter 3. First, the basic concepts involved are introduced and an algorithm for type-2 FRI is proposed. Several illustrative examples are provided to demonstrate the use of this alternative in performing interpolation and extrapolation. Then, a comparison between type-2 FSs and RF sets is discussed. The differences between them are explained in detail. This is followed by an example that illustrates the differences amongst the interpolated results that are obtained by conventional FRI, type-2 FRI, and RF interpolation.

5.1 Implementation with Type-2 Fuzzy Sets

There have been several recent independent developments, e.g., [29, 36, 37, 38, 40, 41, 121, 207] in dealing with FRI using interval type-2 FSs. As an alternative approach, here the preceding framework is implemented using interval type-2 FSs also. In particular, the lower MFs and upper MFs are utilised for describing the uncertainty bounds, with FOU capturing the higher order uncertainty involved.

For rule interpolation, the method calculates an Rep to represent an interval type-2 FS, then derives the interpolated results with transformation techniques. The results are guaranteed to be still interval type-2 FSs to maintain representational consistency. This approach enables interpolation not only for type-2 FSs, but covers the interpolation for conventional FSs as a special case.

Being another realisation of the proposed framework, this implementation with interval type-2 FSs is straightforward and similar to that of using RF sets. For illustration interpolation with rules involving triangular interval type-2 FSs with a single antecedent are outlined here.

Note that a comparative study between RF sets and type-2 FSs will be explained later in Section 5.3. In particular, different calculations for the MFs of RF sets and type-2 FSs will be discussed with an illustrative example in Section 5.3.2.

5.1.1 Representative Values Calculation

The lower and upper Reps and shape diversity factors are respectively calculated following Equations (3.5) and (3.6), such that

$$\mathrm{Rep}(\tilde{A}^{*V})_x = \frac{1}{3}(\tilde{a}_0^{*V} + \tilde{a}_1^{*V} + \tilde{a}_2^{*V}), \quad \mathrm{Rep}(\tilde{A}^{*V})_y = \frac{1}{3}\tilde{H}_{\tilde{A}_1}^{V} \tag{5.1}$$

$$f_{\tilde{A}^*}^{V} = \sqrt{\frac{\sum_{k=0}^{2}(\tilde{a}_k^{*V} - \mathrm{Rep}(\tilde{A}^{*V})_x)^2}{3}} \tag{5.2}$$

where $V = L, U$. The weight factors and overall Reps are respectively computed following Equations (4.5) and (4.6). The calculations for $\tilde{A}_1$ and $\tilde{A}_2$ follow the same procedure.

5.1.2 Closest Rules Selection

The selection of closest rules simply follows Equations (3.9) and (3.10).

5.1.3 Intermediate Rule Construction

The intermediate rule is then constructed following Equations (3.11) - (3.16).

5.1.4 Transformations Implementation

Note that when triangles are used, only three kinds of parameters are needed (s_0^V, $\mathbb{M}_0^V$ and h_1, $V = L, U$). They are calculated as follows:

$$s_0^V = \frac{\tilde{a}_2^{*V} - \tilde{a}_0^{*V}}{\tilde{a}_2^{\prime V} - \tilde{a}_0^{\prime V}} \tag{5.3}$$

$$\mathbb{M}_0^V = \begin{cases} \dfrac{\tilde{a}_0^{*V} - \tilde{a}_0^{\prime\prime V}}{\frac{\tilde{a}_1^{\prime\prime V} - \tilde{a}_0^{\prime\prime V}}{3}} & \text{if } \tilde{a}_0^{*V} \geq \tilde{a}_0^{\prime\prime V} \\ \dfrac{\tilde{a}_0^{*V} - \tilde{a}_0^{\prime\prime V}}{\frac{\tilde{a}_2^{\prime\prime V} - \tilde{a}_1^{\prime\prime V}}{3}} & \text{if } \tilde{a}_0^{\prime\prime V} > \tilde{a}_0^{*V} \end{cases} \tag{5.4}$$

$$h_1 = \frac{\tilde{H}^{*L}_{\tilde{A}_1}}{\tilde{H}^{\prime L}_{\tilde{A}_1}} \tag{5.5}$$

5.1.5 Interpolated Conclusion Derivation

The final interpolated conclusion $\tilde{B}^*$ is obtained using $s^V_{\tilde{B}_0}$, $\mathbb{M}^V_{\tilde{B}_0}$ and $h_{\tilde{B}_1}$, $V = L, U$, where

$$s^V_{\tilde{B}_0} = s_0^V, \;\; \mathbb{M}^V_{\tilde{B}_0} = \mathbb{M}_0^V, \;\; h_{\tilde{B}_1} = h_1 \tag{5.6}$$

5.2 Illustrative Examples

In this section, several examples are used to illustrate the interpolation process, where the observations fall into the rule base “gaps”, using interval type-2 FSs. However, if the observations partially overlap with the rule antecedents and such matches are above a certain confidence level, no interpolation will be required (as conventional compositional rule of inference can then be applied). Otherwise, identical interpolation method can be applied.

5.2.1 Singleton-valued Case

Example 5.1. This case considers one single antecedent variable involving singleton-valued conditions. The involved interval type-2 FSs are listed in Table 5.1.

Table 5.1: Singleton-valued interval type-2 FSs for Example 5.1

Attribute values	$\tilde{A}_1 =< (3,3,3;1),(3,3,3;1) >$
	$\tilde{A}_2 =< (12,13,13.5;0.6),(11,13,14;1) >$
	$\tilde{B}_1 =< (4,4,4;1),(4,4,4;1) >$
	$\tilde{B}_2 =< (10.5,11.5,12;0.5),(10,11.5,13;1) >$
Observation	$\tilde{A}^* =< (6,7,8;0.6),(5,7,9;1) >$

Firstly, the lower and upper Reps, shape diversity factors and weight factors are calculated according to Equations (5.1), (5.2) and (4.5). Secondly, the overall Reps $\text{Rep}(\tilde{A}_1) = 3.333$, $\text{Rep}(\tilde{A}^*) = 7.289$, $\text{Rep}(\tilde{A}_2) = 13.011$ are calculated from Equation (4.6). $\tilde{A}' =< (6.68,7.09,7.29;0.84),(6.27,7.09,7.50;1) >$ and $\tilde{B}' =< (6.66,7.07,7.27;0.80),(6.45,7.07,7.68;1) >$ are then calculated, respectively. Thirdly, the scale rates $s_0^L = 3.26$, $s_0^U = 3.26$, the move ratios $\mathbb{M}_0^L = 0.09$, $\mathbb{M}_0^U = 0.20$ and the height rate $h_1 = 0.72$ in the integrated transformation from $\tilde{A}'$ and $\tilde{A}^*$ are calculated with regard to Equations (5.3), (5.4) and (5.5). Finally, the scale rates, move ratios and height rate are used to transform $\tilde{B}'$ to the interpolated conclusion $\tilde{B}_c^* =< (5.98,7.04,7.98;0.57),(5.27,6.66,9.27;1) >$, as shown in Figure 5.1. It follows that if certain components involved in the given rules are singleton-valued, the interpolated conclusion remains an interval type-2 FS.

5.2.2 Multiple Antecedents Case

Example 5.2. This example concerns an interpolation using rules with multiple antecedent variables. In particular, two rules each involving two antecedents $\tilde{A}_{11} \wedge \tilde{A}_{12} \Rightarrow \tilde{B}_1$, $\tilde{A}_{21} \wedge \tilde{A}_{22} \Rightarrow \tilde{B}_2$ and the observations $\tilde{A}_1^*$, $\tilde{A}_2^*$ are given in order to illustrate the interpolative process to determine the result $\tilde{B}_c^*$. All the conditions are shown in Table 5.2.

In this case, the normalised weight $w'_{\tilde{B}_1}$ for the first antecedent variable is 0.49 and $w'_{\tilde{B}_2}$ for the second is 0.51, they are used to calculate the intermediate rule result $\tilde{B}' =< (6.13,6.87,7.61;0.5),(5.13,6.61,9.10;1) >$ according to Equations (3.15) and (3.16). The average of two lower scale rates (0.68 and 0.89) and the average of two upper scale rates (1.01 and 0.78) are then computed, resulting in $s_0^L = 0.78$ and $s_0^U = 0.90$, forming the combined scale rates. Similarly, the

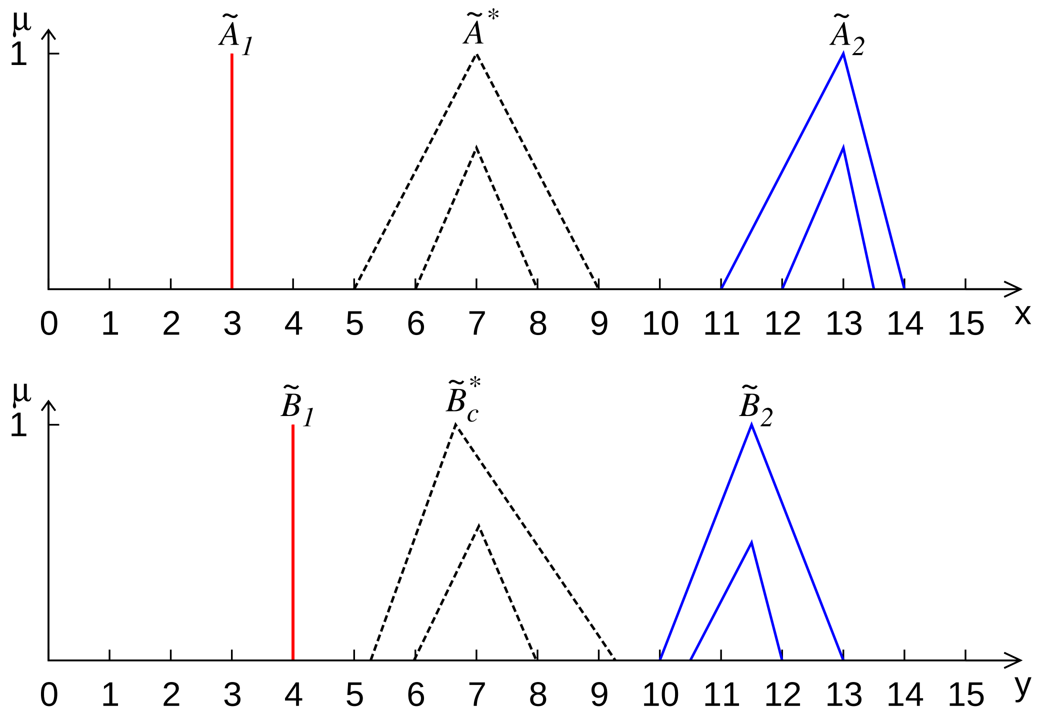


Figure 5.1: Example for a single antecedent case with singleton-valued conditions

Table 5.2: Multiple antecedents case for Example 5.2

Attribute values	$\tilde{A}_{11} =< (1, 3.5, 4; 0.7), (0, 4, 5; 1) >$
	$\tilde{A}_{21} =< (12, 13, 13.5; 0.7), (11, 13, 14; 1) >$
	$\tilde{A}_{12} =< (1.5, 3.5, 4.5; 0.6), (0, 3.5, 6; 1) >$
	$\tilde{A}_{22} =< (12.5, 13.5, 14; 0.6), (11.5, 13.5, 14.5; 1) >$
	$\tilde{B}_1 =< (1, 2, 3; 0.5), (0, 2, 5; 1) >$
	$\tilde{B}_2 =< (11, 11.5, 12; 0.5), (10, 11, 13; 1) >$
Observation	$\tilde{A}_1^* =< (7.5, 8, 9; 0.7), (6, 8, 10; 1) >$
	$\tilde{A}_2^* =< (7.5, 8, 9.5; 0.6), (6.5, 8, 10; 1) >$

combined move ratios $\mathbb{M}_0^L = 0.79$ and $\mathbb{M}_0^U = 0.27$ are calculated from two lower move ratios (0.84 and 0.75) and two upper move ratios (0.28 and 0.26). These, together with the combined height rate, namely the average $h_1 = 1$ of the two height rates (1 and 1), are employed to transform $\tilde{B}'$ to achieve the final result $\tilde{B}_c^* =< (6.47, 6.61, 7.57; 0.5), (5.41, 6.37, 9.04; 1) >$, with $\delta_0 = 1.20$, $\delta_1 = 1.04$ and $\delta_2 = 0.84$, as shown in Figure 5.2.

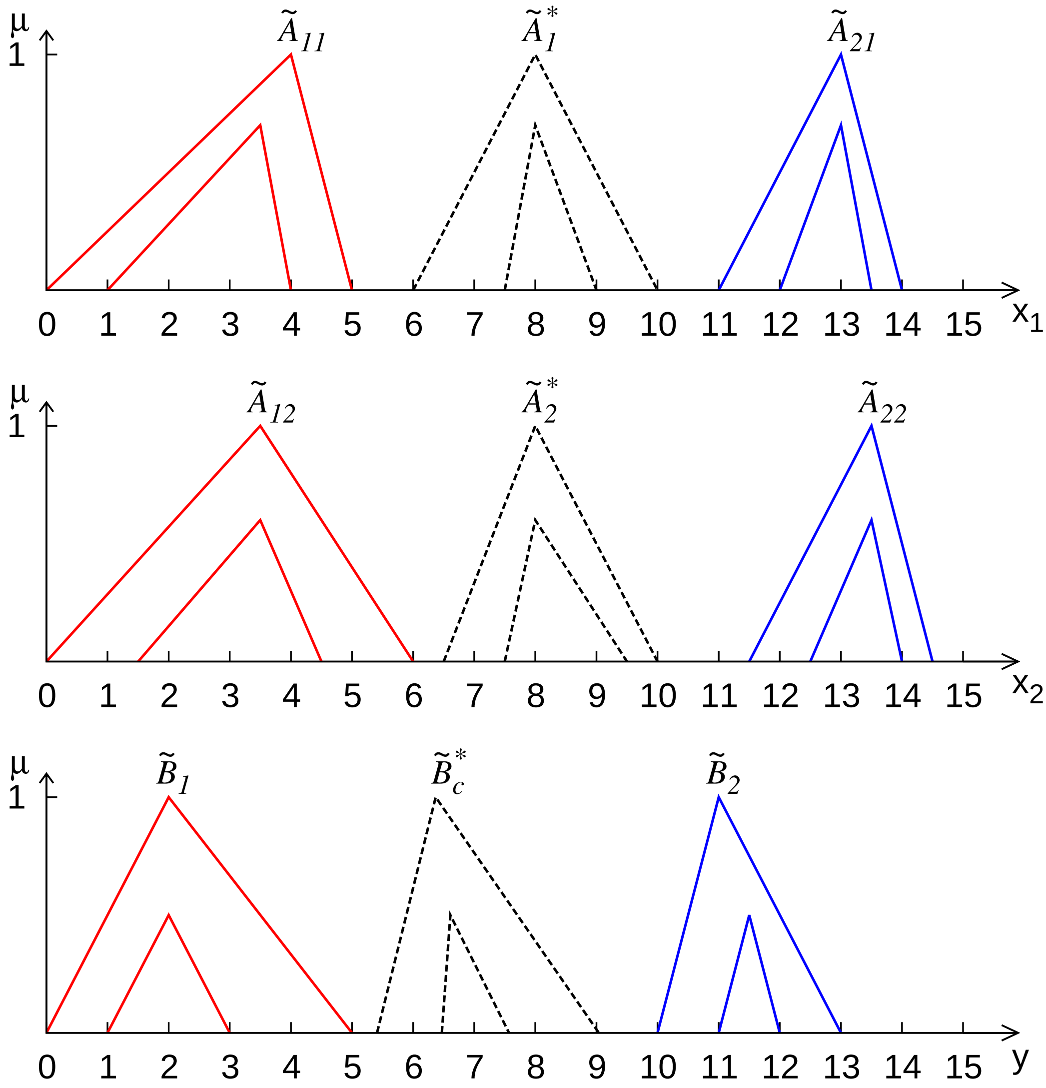


Figure 5.2: Example for a multiple antecedents case

5.2.3 Multiple Rules Case

Example 5.3. This example considers a general multiple multi-antecedent rules case, where four rules with three antecedents are selected as the neighbouring rules. In particular, two antecedents of a given observation are located between these rules, whereas one antecedent is located beyond them. This relates to a hybrid case, including interpolation and extrapolation. All the involved FSs are listed in Table 5.3.

For the antecedent dimension x_1, the normalised weight of Rule 1 is 0.37. Another two weights for x_2 and x_3 can be computed in the same way, resulting in 0.12 and

Table 5.3: Interpolation and extrapolation hybrid

	Antecedents	Consequence
	$\tilde{A}_{11} =< (1, 2, 3; 0.7), (0, 1.5, 4; 1) >$	
Rule 1	$\tilde{A}_{12} =< (2, 3, 4; 0.5), (1, 2.5, 5; 1) >$	$\tilde{B}_1 =< (1, 1.5, 2.5; 0.6), (0, 1.5, 3.5; 1) >$
	$\tilde{A}_{13} =< (7, 8, 9; 0.6), (6, 7.5, 10; 1) >$	
	$\tilde{A}_{21} =< (16, 17.5, 19; 0.7), (15, 17.5, 20; 1) >$	
Rule 2	$\tilde{A}_{22} =< (11.5, 12.5, 13; 0.5), (10, 13, 15; 1) >$	$\tilde{B}_2 =< (21.5, 23, 23.5; 0.6), (20.5, 23, 24; 1) >$
	$\tilde{A}_{23} =< (21, 22, 23; 0.6), (20, 22, 23.5; 1) >$	
	$\tilde{A}_{31} =< (21.5, 22.5, 24; 0.7), (20, 22, 25; 1) >$	
Rule 3	$\tilde{A}_{32} =< (6.5, 7, 8.5; 0.5), (5.5, 7.5, 9; 1) >$	$\tilde{B}_3 =< (16.5, 17, 18; 0.6), (15, 17, 20; 1) >$
	$\tilde{A}_{33} =< (12, 12.5, 14; 0.6), (11, 13, 15; 1) >$	
	$\tilde{A}_{41} =< (11.5, 12, 13; 0.7), (10.5, 12, 14; 1) >$	
Rule 4	$\tilde{A}_{42} =< (22, 23, 23.5; 0.5), (21, 22.5, 24; 1) >$	$\tilde{B}_4 =< (12, 13, 14; 0.6), (11.5, 13.5, 14.5; 1) >$
	$\tilde{A}_{43} =< (17, 18.5, 19; 0.6), (16, 18.5, 19.5; 1) >$	
	$\tilde{A}_1^* =< (6, 6.5, 7.5, ; 0.7), (5, 6.5, 9; 1) >$	
Observation	$\tilde{A}_2^* =< (16.5, 18, 19.5; 0.5), (15, 18, 20; 1) >$	
	$\tilde{A}_3^* =< (1.5, 2.5, 4; 0.6), (0.5, 3, 5; 1) >$	

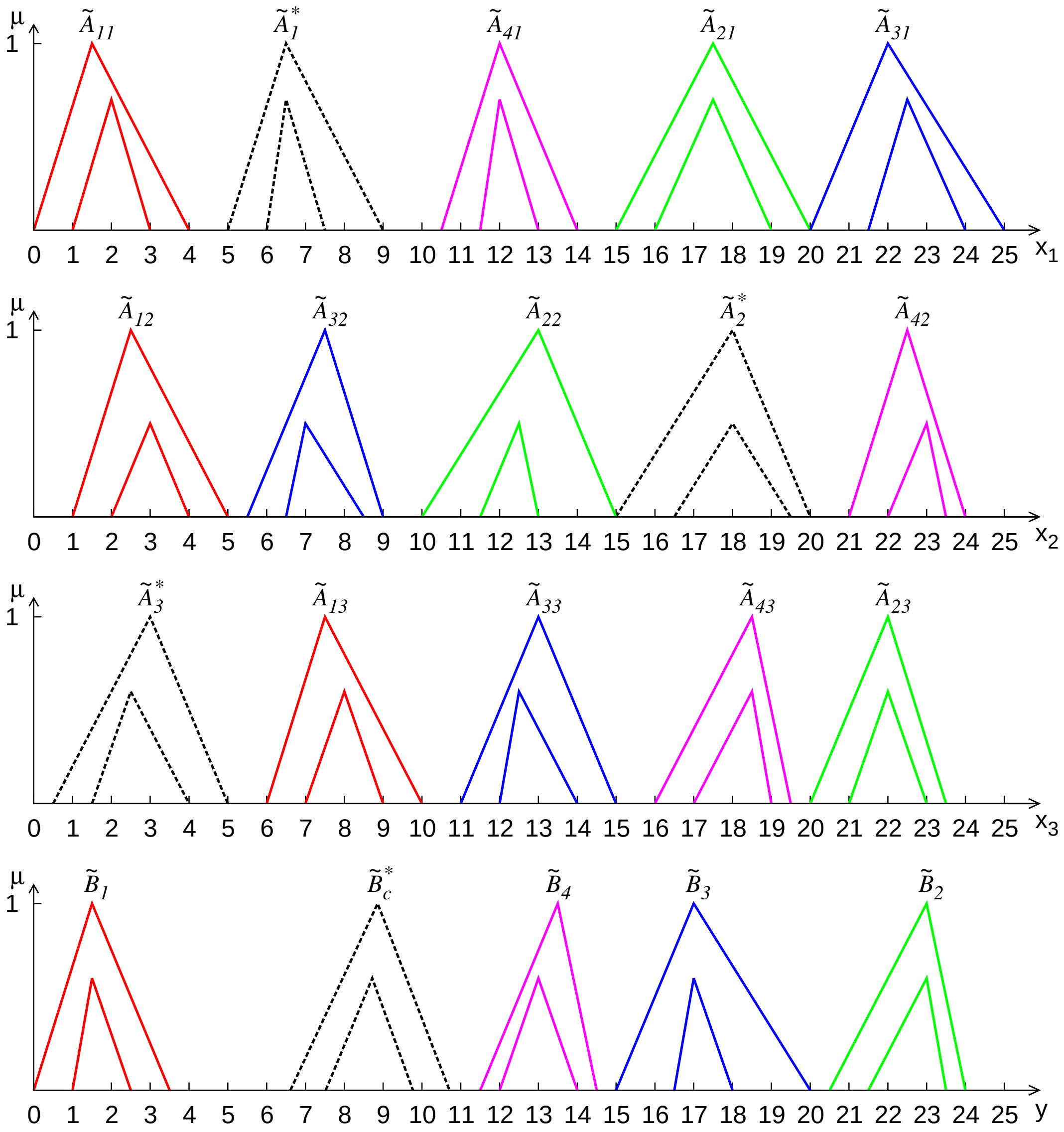


Figure 5.3: Example for a multiple rules case

0.47. In this case, the weight that Rule 1 contributes to the intermediate rule is 0.32. The intermediate rule can then be constructed by implementing the weight calculation on each rule. After this, the scale rates $s_0^L = 1.27$ and $s_0^U = 1.14$ are generated according to Equation 5.3. Similarly, the move ratios $\mathbb{M}_0^L = -0.09$ and $\mathbb{M}_0^U = -0.02$ are obtained with respect to Equation 5.4, while the height rate $h_1 = 1$ is obtained with respect to Equation 5.5. Finally, these parameters are utilised to achieve the conclusion $\tilde{B}_c^* = <(7.52, 8.72, 9.77; 0.6), (6.61, 8.86, 10.71; 1)>$, as shown in Figure 5.3. It can be seen that the interpolation and extrapolation hybrid performs well.

5.2.4 Non-general Case

Example 5.4. This example demonstrates the use of the proposed approach involving only type-1 FSs. All the terms are listed in Table 5.4, where the type-1 FSs are still represented in the form of interval type-2 FSs.

Table 5.4: Involved interval type-2 FSs for Example 5.4

Attribute values	$\tilde{A}_1 =< (0,5,6;1),(0,5,6;1) >$
	$\tilde{A}_2 =< (11,13,14;1),(11,13,14;1) >$
	$\tilde{B}_1 =< (0,2,4;1),(0,2,4;1) >$
	$\tilde{B}_2 =< (10,11,13;1),(10,11,13;1) >$
Observation	$\tilde{A}^* =< (7,8,9;1),(7,8,9;1) >$

The final interpolated conclusion $\tilde{B}_c^* =< (5.83,6.26,7.38;1),(5.83,6.26,7.38;1) >$ can be derived, as shown in Figure 5.4. The same result can be found in [76]. This implies the compatibility to the original T-FRI. However, the proof is omitted here in order to avoid repetition. Similar description can be found in Section 4.5.

5.3 Comparison to Rough-Fuzzy-based Implementation

A comparative study is provided in this section. Conceptual comparison between type-2 FSs and RF sets is given first. Experimental comparison is then presented to show the differences amongst the interpolated results obtained by different means.

5.3.1 Type-2 Fuzzy Sets vs. Rough-Fuzzy Sets

A general type-2 FS replaces the crisp valued membership grades of a type-1 FS with those of FS(s). For each value of the primary variable, the membership is a function (the secondary MF), whose domain (the primary membership) is in the interval $[0,1]$, and whose range (secondary membership grades) may also be in the interval $[0,1]$. Therefore, the MF of a general type-2 FS is three-dimensional, and it is the new third-dimension that provides an additional degree of freedom that makes it possible to directly handle higher order uncertainty [133]. Unfortunately,

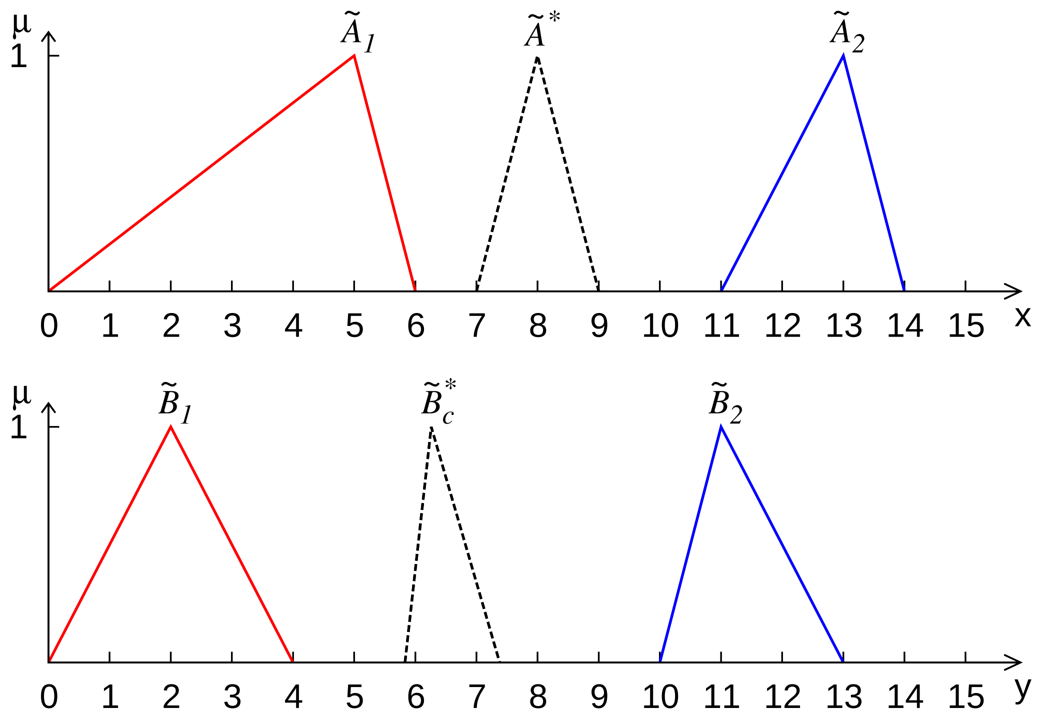


Figure 5.4: Example for a non-general (type-1 FSs) case

the secondary MF is still described using crisp values. The task of determining these values becomes a new dilemma, which has it roots in the same dilemma as that of using conventional FSs. This dilemma still exists even when using type-n FSs ($n > 2$). This seems to be a recursive problem that gives rise to other problems, while the original problem is at best reduced but not removed.

Type-2 FSs also include a specific group of FSs that are referred to as interval type-2 FSs [136], where all the values of the secondary membership grades are the same. In this case, the third-dimension is no longer needed because it conveys no new information. Although general type-2 FSs have one extra degree of freedom than interval type-2 FSs, it is not yet known how to best choose their secondary MFs [135]. The third-dimension is therefore often ignored in order to reduce the computational complexity, however this may result in a non-intuitive lack of the desirable degree of freedom.

Note that uncertainty in the primary memberships of a general type-2 FS consists of a region FOU that is bounded by the upper and lower MFs. Because the secondary membership grades convey no new information, the FOU is a complete description of an interval type-2 FS [136]. Thus, it is only this single region that is used to describe uncertainty in interval type-2 FSs.

In contrast, an RF set describes the representation of uncertain FS MFs through the use of RF approximations. The upper and lower approximations are derived by the given FSs, which are known to belong exactly to a given concept. In particular, the UA is defined by the union of all elements of the given sets, while the LA is defined by the intersection of all the given sets. Hence, the calculus of RF sets can be applied in a purely data-driven manner, no additional subjective definitions or thresholds are needed. Importantly, the approximation MF of an RF set is two-dimensional because both approximations are directly derived from the MFs of FSs of a two-dimensional space. Since there is no need for the third dimension, no unknown uncertain information is added. Thus, the computational complexity can be lower than that required to deal with general type-2 FSs.

The concept of RF sets is based on the definition of RSs. An RF set is constructed by the upper and lower approximation MFs. The UA indicates the individual region that is given by at least one person (uncertainty + certainty), and the LA indicates the common region that is agreed by all persons (certainty). Moreover, the boundary region, bounded by the two approximations, indicates the region that can possibly, but not always certainly, be partitioned in this way (uncertainty). If the boundary region is an empty set, namely, the LA coincides with the UA, then the RF set degenerates to a conventional FS. In this case, all uncertainty disappears. The area of the boundary region determines the degree of uncertainty involved in such an RF set. That is, the closer the shapes of the lower and upper approximations, the lower the uncertainty of the set. Therefore, two regions (three if the boundary region which can be determined by the UA and LA is considered as a separate one) are utilised to describe uncertainty in RF sets.

The UA plays the same role as the interval in interval-valued FSs [19, 53, 172]. They both involve all possibilities, but it is not possible to tell which values are given by all people who interpret an uncertain concept. This important information is lost as a result of interval-valued FSs. Interval type-2 FSs share the same problem, as the FOU is similar to the interval, because the FOU is the union of all primary memberships (intervals) [136]. The third-dimension of general type-2 FSs has the ability to provide this information, but as aforementioned, it is not known how this can be computationally implemented in an efficient way. Furthermore, the third-dimension is set to value 1 in order to reduce the computational complexity. In contrast, this information can be easily captured by means of the LA of RF sets. The LA indicates a region that is definitely covered by the MF defining an uncertain concept. That is, if a conventional FS is used to describe an uncertain concept, the representation of the FS MF must contain the region of the LA. Similarly, the representation should not go beyond the region of the UA.

As a consequence, RF sets and type-2 FSs are two different extensions of conventional FSs. The differences between them are listed in Tables 5.5 and 5.6. Also, an example is provided to show the generation process below.

Table 5.5: Comparison with general type-2 FSs

	General type-2 FSs	RF sets
Foundation	Fuzzy-fuzzy sets	RF approximations
Definition spaces	Three-dimensional	Two-dimensional
Computational complexity	High, additional definitions required for third-dimension	Low, data-driven, no additional subjective definitions required

Table 5.6: Comparison with interval type-2 FSs

	Interval type-2 FSs	RF sets
Foundation	General type-2 FSs ignored the third-dimension	RF approximations
Definition spaces	Two-dimensional	Two-dimensional
Computational complexity	Low, but may result in a lack of additional degrees of freedom	Low, data-driven
Regions used for uncertainty describing	One: FOU	Two: LA and UA

5.3.2 Illustrative Example and Discussion

Suppose that the concept of interest is *eye contact* [133], denoted by x with x belonging to the intensity range of $[0, 10]$, where 0 indicates no eye contact and 10 represents maximum eye contact. One of the terms that may characterise the amount of perceived *eye contact* is “some eye contact”. Suppose that 10 people are surveyed, and are asked to locate the distribution for *some eye contact* on the scale 0-10, as listed in Table 5.7. It is reasonable to assume that the same results are not obtained from 10 individuals. Instead, different MFs to depict such a fuzzy concept are likely to be given. The question is how such diverse information about the underlying concept may be captured and described conjunctively.

Table 5.7: Data from 10 people

10 individuals	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
Preferred degrees	3.5	5	4.5	5.5	4	5	6	4.5	6	5
Acceptable ranges	1-6	1-9	1.5-7.5	1.5-9.5	2-6	2-8	2.5-9.5	3-6	3-9	3.5-6.5

One approach might be to construct a type-1 FS with triangular-shaped MF whose base endpoints (on the x-axis) are at the two average endpoint values and whose apex is midway between these two endpoints. Such a conventional FS is derived in two dimensions, as shown in Figure 5.5. However, this approach completely ignores the uncertainty associated with the different results.

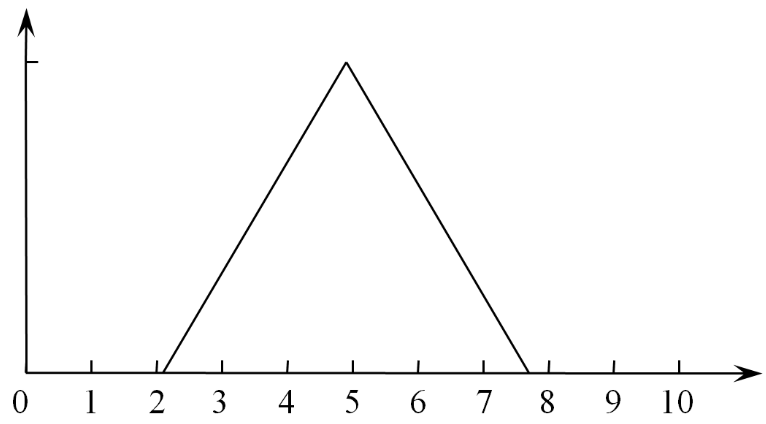


Figure 5.5: MF of a conventional FS

An alternative approach is to make use of the average endpoint values and the standard deviation of each endpoint to establish an uncertainty interval about each average endpoint value. By doing this, for each x, the MF is no longer a single value, instead, it is itself a function. A general type-2 FS with three-dimensional MF is created this way for all x, as shown in Figure 5.6, where two endpoints have two uncertainty intervals associated with them, and the apex point is assumed to have a full certainty value. However, in general, the apex point can also have an uncertainty interval associated with it, which cannot be modelled in this approach.

A third approach is to calculate the union and intersection of all the given FSs to decide the common and individual regions about the uncertain concept of interest.

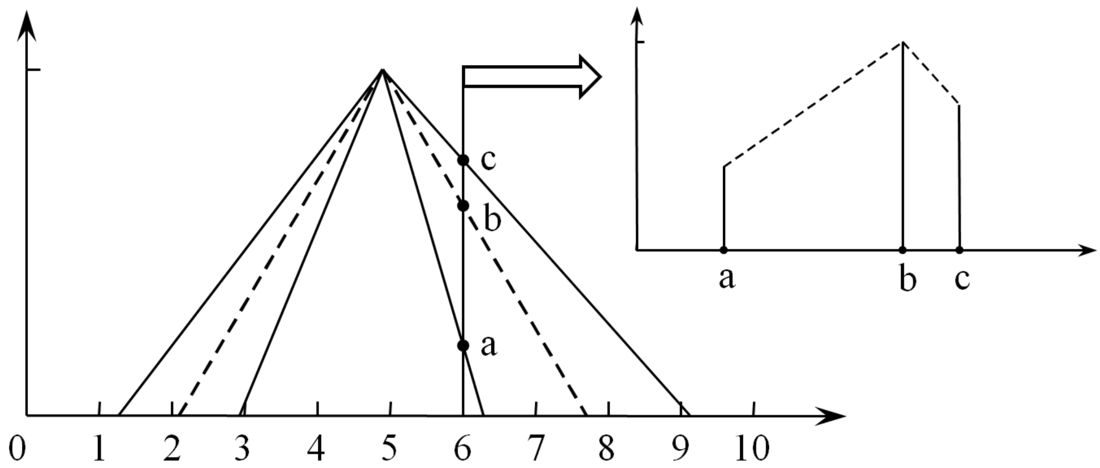


Figure 5.6: MF of a general type-2 FS

For each x, the MF of the union of 10 sets is defined as the maximum of the 10 individual MFs. Similarly, the membership of the intersection of 10 sets is defined as the minimum of the 10 individual MFs. The resulting two MFs are together referred to as an RF set. Such an RF set is defined in two-dimensional space, as shown in Figure 5.7.

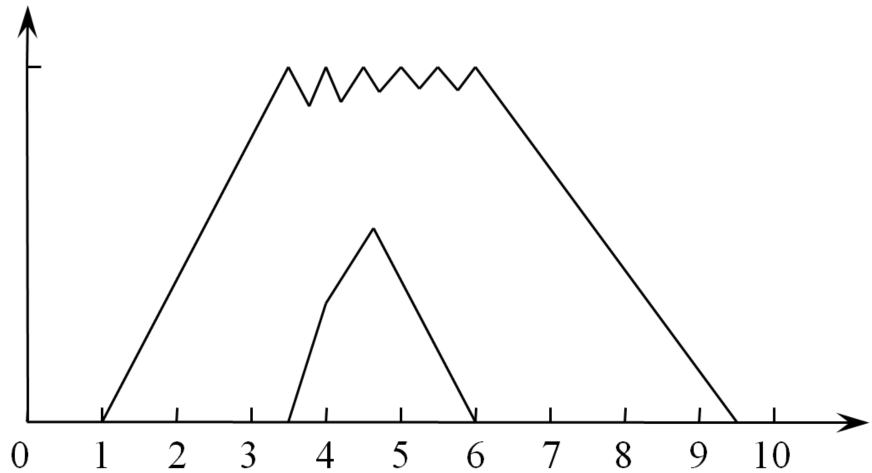


Figure 5.7: MF of an RF set

5.3.3 Illustrative Example for Interpolation with Different Representations

Given the above obtained conventional representation, type-2 representation, and RF representation, an example is presented in this subsection in order to further evaluate the differences between their use in performing interpolation. Three neighbouring rules with each having three antecedent variables are considered. Each of the three representations is used to express the observation in the antecedent. This is feasible because the preceding framework provides an unified realisation for this experiment.

All the involving rules are listed in Table 5.8. Note that the LA and UA of the RF representation are simulated using triangular MFs in order to make consistent with the other two representations (although this will lead to the loss of certain information for the resultant UA). The three interpolated processes and results are shown in Figures 5.8, 5.9, and 5.10, respectively.

Table 5.8: Three neighbouring rules

Rule 1	$\tilde{A}_{11} =< (10.5, 12, 14; 1), (10.5, 12, 14; 1) >$
	$\tilde{A}_{12} =< (21, 22.5, 24; 1), (21, 22.5, 24; 1) >$
	$\tilde{A}_{13} =< (16, 18.5, 19.5; 1), (16, 18.5, 19.5; 1) >$
	$\tilde{B}_{1} =< (11, 13.5, 15.5; 1), (11, 13.5, 15.5; 1) >$
Rule 2	$\tilde{A}_{21} =< (20, 22, 25; 1), (20, 22, 25; 1) >$
	$\tilde{A}_{22} =< (15, 18, 20; 1), (15, 18, 20; 1) >$
	$\tilde{A}_{23} =< (11, 13, 15; 1), (11, 13, 15; 1) >$
	$\tilde{B}_{2} =< (16, 17.5, 19.5; 1), (16, 17.5, 19.5; 1) >$
Rule 3	$\tilde{A}_{31} =< (15, 17.5, 20; 1), (15, 17.5, 20; 1) >$
	$\tilde{A}_{32} =< (10, 13, 15; 1), (10, 13, 15; 1) >$
	$\tilde{A}_{33} =< (20, 22, 23.5; 1), (20, 22, 23.5; 1) >$
	$\tilde{B}_{3} =< (20.5, 23, 24; 1), (20.5, 23, 24; 1) >$

It is obvious that these results are different in detail. That is to say, different representations lead to different interpolated conclusions. However, the results achieved by type-2-based and RF-based are similar, and both cover the conventional result as their specific case. In particular, the left (right) endpoint of the conventional representation is located between the two left (right) endpoints of the type-2 and RF representations. This is reasonable as the uncertain information is reserved in the type-2 and RF representations and therefore reflected in the interpolated results. Intuitively, more useful information, which is involved in the interpolation process,

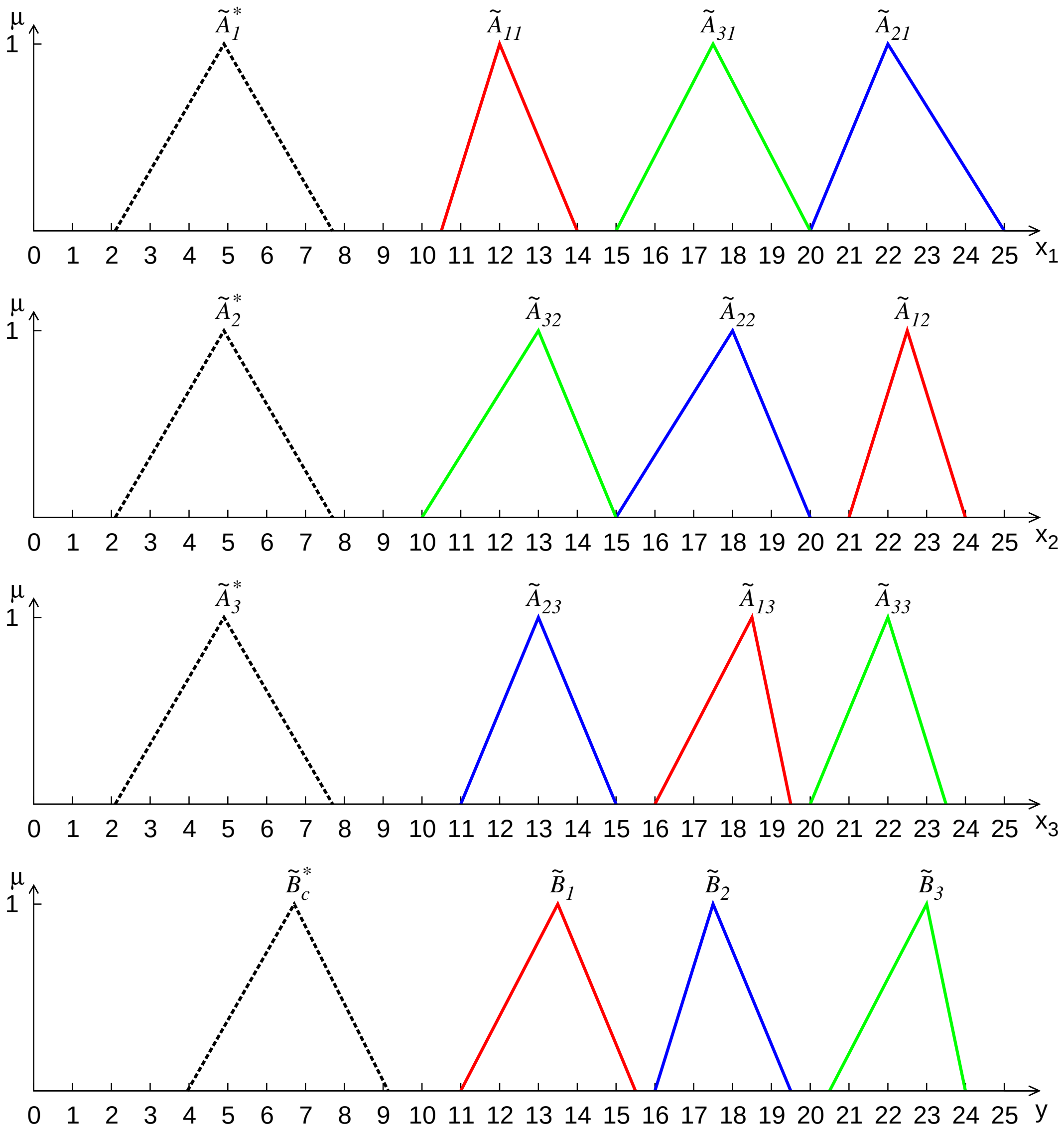


Figure 5.8: Interpolated result of the conventional representation

may obtain better interpolated results. This will be evaluated and discussed with application to realistic decision making problems in Chapter 6.

5.4 Summary

This chapter has presented an alternative implementation of the proposed framework using type-2 FSs. Thanks to the generality of the framework, the type-2 FRI method can be built in a straightforward manner having developed the RF version. Examples have been presented to demonstrate this successful alternative implementation,

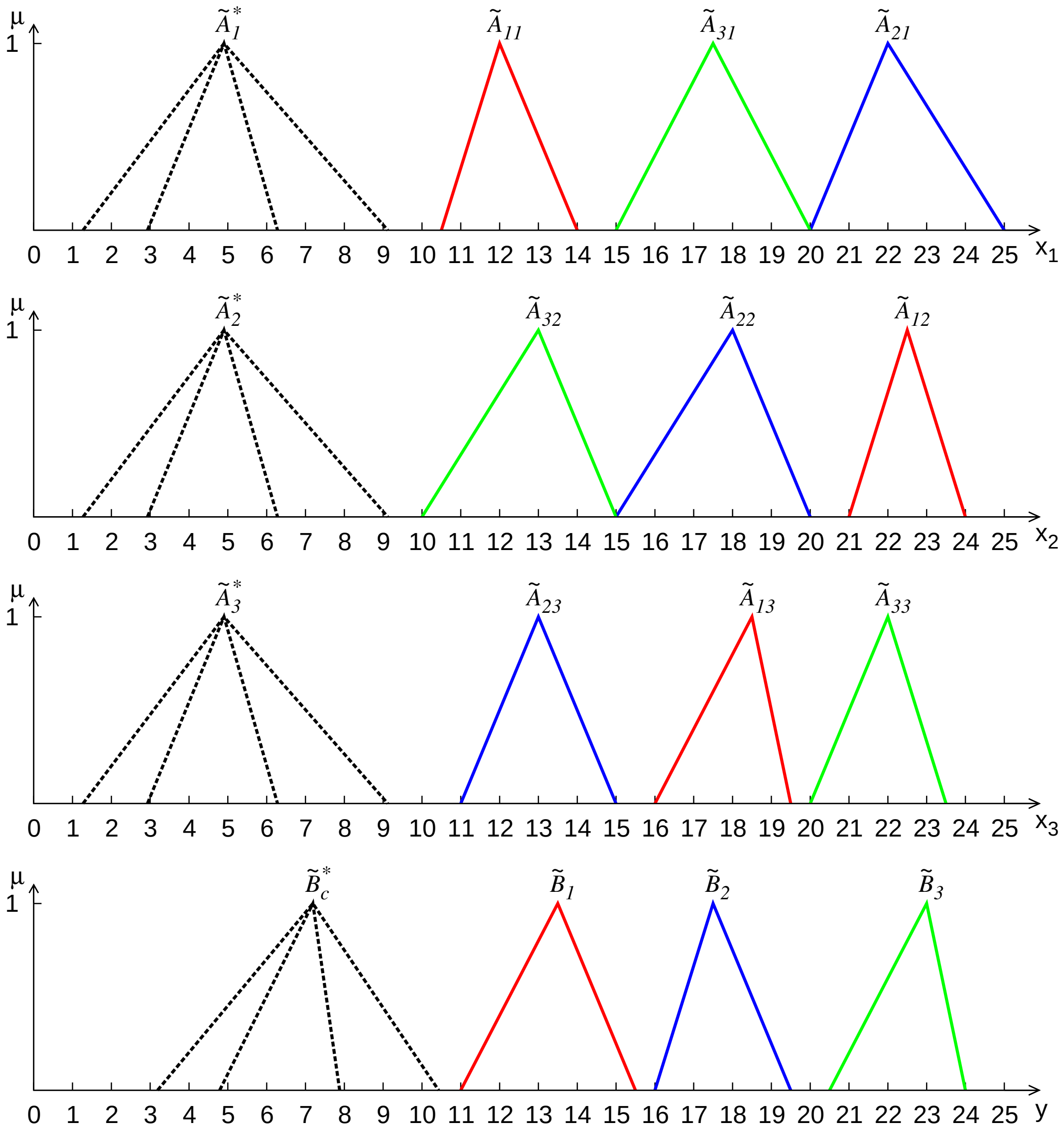


Figure 5.9: Interpolated result of the type-2 representation

including singleton-valued, multiple antecedents, multiple rules, and non-general cases. These examples indicate that the type-2 FRI is also a useful extension of the existing type-1 FRI. A comparison between type-2 FSs and RF sets has been provided. Discussion concerning general type-2 FSs and interval type-2 FSs has also been given, supported with an illustrative example.

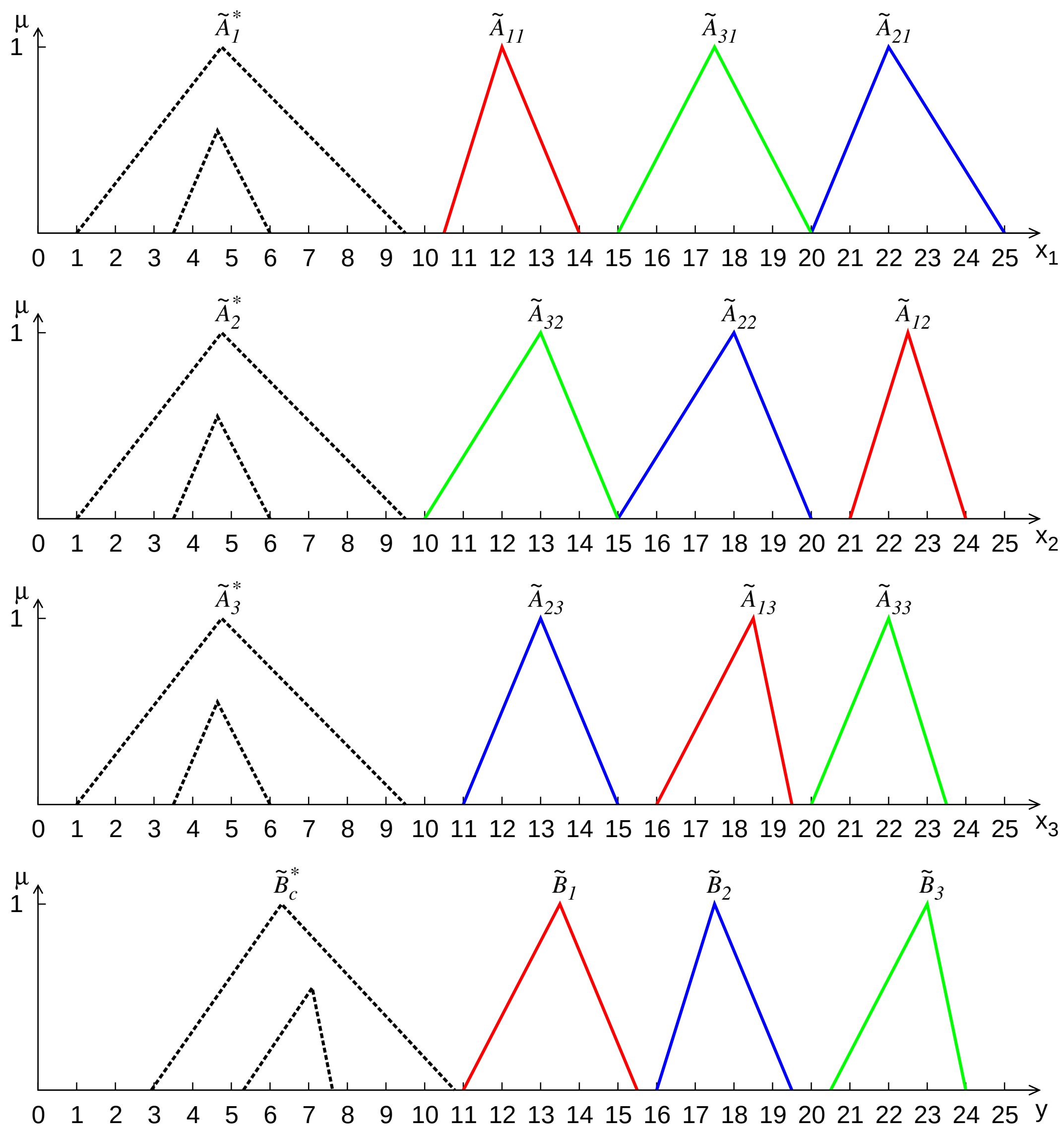


Figure 5.10: Interpolated result of the RF representation

Chapter 6

Higher Order Fuzzy Rule Interpolation: Evaluations

EXPERTS have always attempted to model how environmental change may influence disease burden so that they can predict the relevant disease rate. However, the models built for this are often very complicated and due to typically very sophisticated situations, usually result in a sparse knowledge or rule base. Moreover, different experts may have different kinds of expertise, leading to similar but different expert rules and observations. Therefore, such problems provide a potentially suitable testbed for the framework proposed earlier and their implementations. To further evaluate the work, the UCI datasets [3] are also utilised for verifying the efficacy of the proposed RF approach.

6.1 Application to Diarrhoeal Disease Prediction

In this section, the effectiveness of the proposed framework in improving the robustness of FRI is demonstrated by a practical application of predicting diarrhoeal disease rates in remote villages.

6.1.1 Problem Overview

Environmental change influences disease burden [42, 144]. Intensive studies have been made in an effort to identify logical relationships underlying such influences so that the consequences of a certain environmental change may be predicted. This is of significant importance in the assessment of potential impact of such changes upon the environment and society, before the starting of any large-scale infrastructure projects.

One particular application problem in this area has recently been investigated in [206, 208], which is based on the study of [58]. It addresses the issue of measuring how the construction of a new road or railway in a previously roadless area may affect the epidemiology of infectious diseases in northern coastal Ecuador. A predictive model has been built where many involved factors are not linearly related, but interact with each other in a grid network. Addressing this application problem, an illustrative example is presented here to show the working of the higher order FRI, especially that of using RF sets and type-2 FSs in the implementation. The original problem of [206] is simplified such that all the studied factors are linearly connected. The resulting simpler causal model is shown in Figure 6.1.

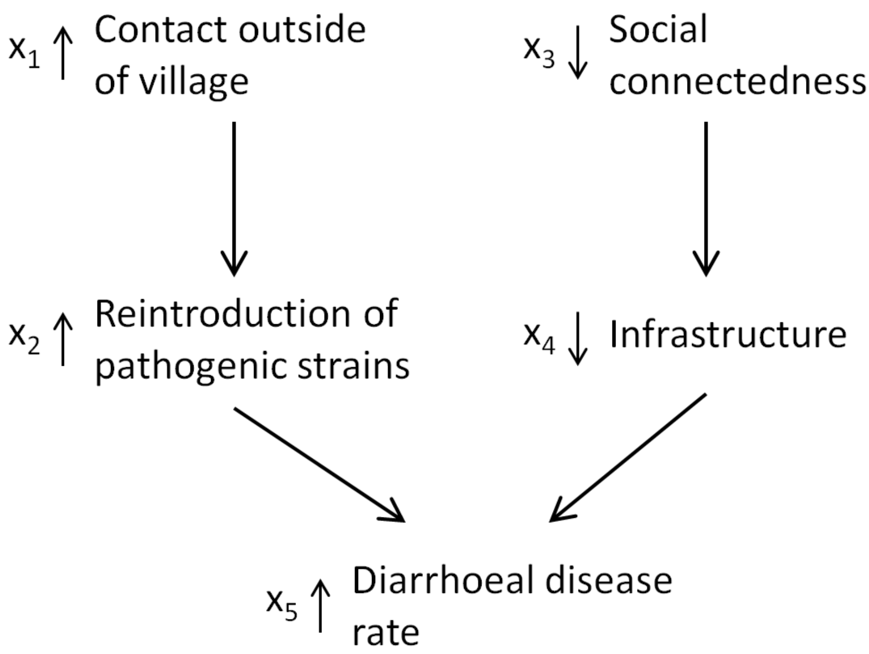


Figure 6.1: Causal diagram of a simplified application problem

This causal diagram shows that the diarrhoeal disease rate of a remote village is directly affected by two factors. First, low social connectedness tends to failure in creating adequate water and sanitation infrastructure because the residents are unlikely to know one another well and share social norms [7, 65], thereby usually resulting in a high diarrhoeal disease rate. Second, more frequent contact between the residents within a village and those outside tends to increase the rate of introduction of pathogens, thereby also raising the diarrhoeal disease rate.

All factors considered in this example are represented as system variables and each relation between two directly connected factors is represented as a rule associating the relevant variables. In summary, there are five variables in the problem: contact outside of the village, reintroduction of pathogenic strains, social connectedness, hygiene and sanitation infrastructure, and infections disease rate, denoted as $x_1, \ldots, x_5$, respectively. Note that different variables are defined on different domains. To simplify knowledge representation, variable domains are mapped onto the real line and normalised.

In order to evaluate the final disease rate, a group of experts are selected to express their views on each factor. Suppose that the opinions from six experts, denoted as $T_1, \ldots, T_6$, in the group are shown in Figure 6.2, where subsets of rules (one subset per causal implication): $A \to B$, $C \to D$ and $B \wedge D \to E$ are established by the experts with each supported by two of them.

6.1.2 Experimentation and Discussion

Given different expert rules and observations, one way to resolve the problem might be to use a conventional FRI approach, say T-FRI to implement required interpolation separately. Suppose that two pairs of expert rules are contained in a sub-rule base: $A_1 \to B_1$ and $A_2 \to B_2$, where $A_{11} \to B_{11}$ and $A_{21} \to B_{21}$ are provided by expert T_1, while $A_{12} \to B_{12}$ and $A_{22} \to B_{22}$ are provided by expert T_2. Note that $A = x_1$, $B = x_2$, $C = x_3$, $D = x_4$ and $E = x_5$. Presented with two observations A_1^* and A_2^*, the interpolated result by the use of T-FRI is a set which contains 4 elements. The computation with respect to the remainder of the subsets of rules follows the same procedure, resulting in a consequence set of 32 interpolated results, as listed in Table 6.1.

Note that the cardinality of the set of interpolated consequent results increases rapidly along with the increase of the cardinality of rule subsets and the number of observations. Suppose that there are m_1 rules in $A \to B$, m_2 rules in $C \to D$, m_3 rules in $B \wedge D \to E$, n_1 observations in A^*, and n_2 observations in C^*. Then the cardinality of the consequence set is $|m_1 n_1 m_2 n_2 m_3|$. This not only leads to

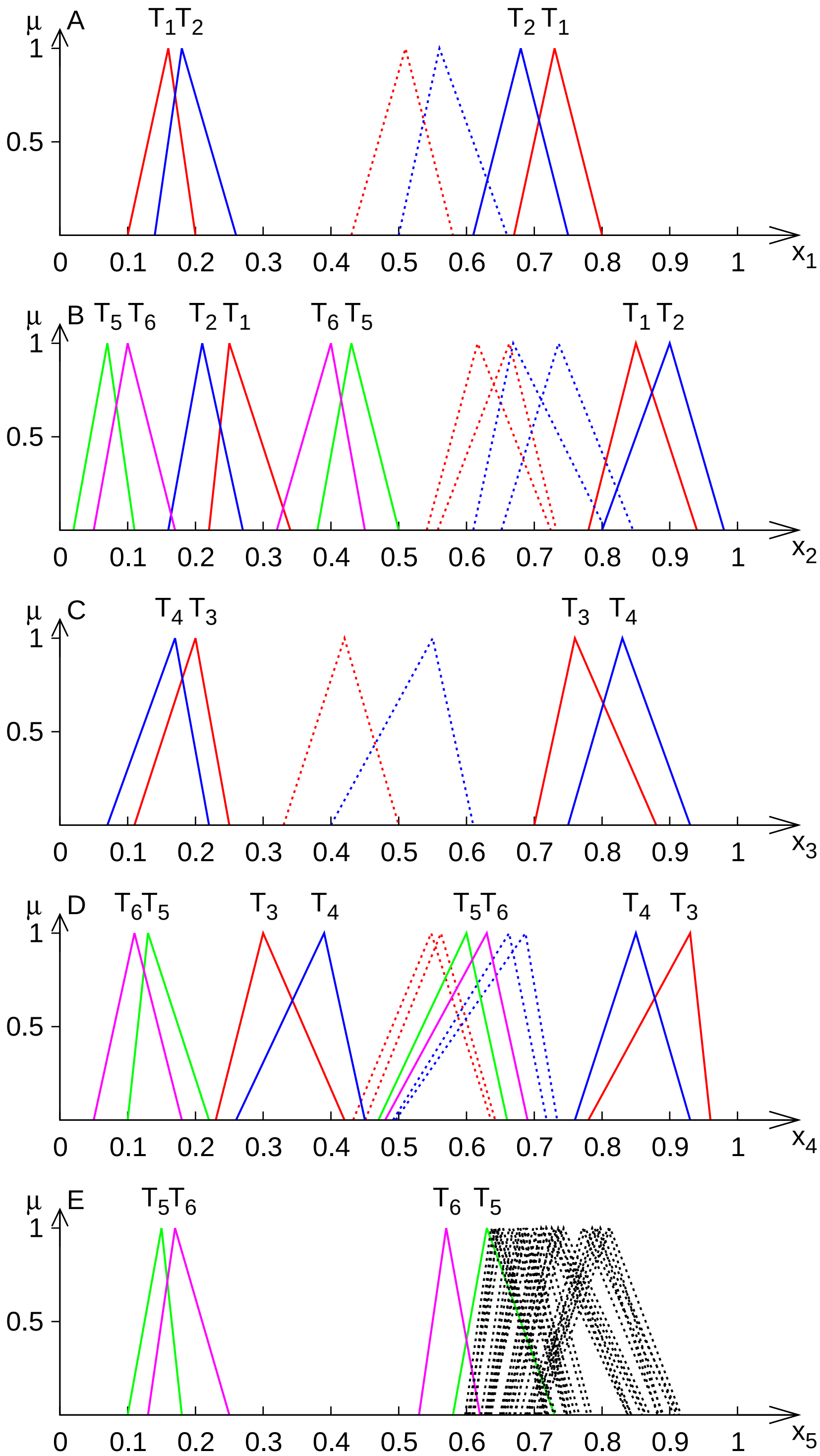


Figure 6.2: Interpolated results from conventional FRI

Table 6.1: Interpolated results

	Values		Values		Values		Values
E_1^*	(0.649,0.710,0.838)	E_2^*	(0.658,0.718,0.843)	E_3^*	(0.689,0.785,0.891)	E_4^*	(0.686,0.771,0.884)
E_5^*	(0.654,0.730,0.838)	E_6^*	(0.664,0.738,0.843)	E_7^*	(0.692,0.809,0.889)	E_8^*	(0.690,0.794,0.882)
E_9^*	(0.663,0.718,0.859)	E_{10}^*	(0.672,0.726,0.864)	E_{11}^*	(0.705,0.789,0.913)	E_{12}^*	(0.702,0.775,0.906)
E_{13}^*	(0.670,0.735,0.865)	E_{14}^*	(0.679,0.743,0.870)	E_{15}^*	(0.710,0.812,0.917)	E_{16}^*	(0.707,0.798,0.910)
E_{17}^*	(0.598,0.636,0.717)	E_{18}^*	(0.605,0.642,0.720)	E_{19}^*	(0.629,0.684,0.754)	E_{20}^*	(0.627,0.676,0.748)
E_{21}^*	(0.602,0.648,0.718)	E_{22}^*	(0.609,0.654,0.721)	E_{23}^*	(0.631,0.699,0.754)	E_{24}^*	(0.630,0.689,0.749)
E_{25}^*	(0.605,0.640,0.729)	E_{26}^*	(0.611,0.646,0.732)	E_{27}^*	(0.636,0.687,0.767)	E_{28}^*	(0.634,0.679,0.761)
E_{29}^*	(0.622,0.663,0.746)	E_{30}^*	(0.628,0.669,0.749)	E_{31}^*	(0.654,0.713,0.784)	E_{32}^*	(0.652,0.703,0.778)

difficulty in interpreting the results, but also causes high computational complexity. As outlined previously, the first step of interpolation requires the computation of the closest rules from a given rule base. A distance measure needs to be calculated in order to estimate the proximity between each rule antecedent and the observation. This implies a time complexity of $O(xyz)$, where x is the number of observations to be interpolated, y is the number of antecedent variables, and z is the number of fuzzy rules involved in a rule subset. From this, the time complexities for the rule subsets depicting the relations $A \to B$, $C \to D$ and $B \wedge D \to E$ are $O(m_1 n_1)$, $O(m_2 n_2)$ and $O(m_1 n_1 m_2 n_2 m_3)$, respectively. Besides, this leads to difficulty in determining the final result. For example, consider two interpolated results $E^*_{15} = (0.710, 0.812, 0.917)$ and $E^*_{17} = (0.598, 0.636, 0.717)$. Using the method in [147], the similarity between these two FSs is 0.002. In this case, it is difficult to make a choice whilst they are almost completely different conclusions.

Fortunately, the proposed RF approach can be applied without suffering from this difficulty. All the uncertain relations can be captured using RF sets and the conclusion can be derived by RF interpolation. The interpolated results following the present work are illustrated in Figure 6.3. These results reflect the distribution of those results shown in Figure 6.2. In particular, the shape of the resultant RF set is similar to the shape distribution of those 32 interpolated sets, whereas the computational complexity of the former is much lower than that of the latter. This can be noticed by comparing the calculated time complexities of the former, which are $O(m_1)$, $O(m_2)$ and $O(m_3)$, respectively. It is obvious that the reduction in computation complexity is significant, especially when the number of observations becomes large. In addition, since a majority of the 32 results are closer to the right rule, the resultant RF set is also closer to it. The reason for this is that the RF sets are defined based on both common and personal information.

Similarly, the proposed type-2 implementation can also be used for this problem. The result is shown in Figure 6.4. It can be seen that the two resultant sets have similar locations. This implies that both approaches are effective in finding approximate solutions for this problem. The difference between them is the shape of the interpolated results. In particular, the LA area of the RF set is smaller. As indicated previously, E^*_{15} and E^*_{17} have less overlapping. This situation is reflected by the RF approach and therefore results in the smaller LA. Unfortunately, since the mean and the standard deviation cannot represent the overlapping, this is not showed in the type-2 approach. Nevertheless, both results do seem to reveal that both implementations perform reasonably well.

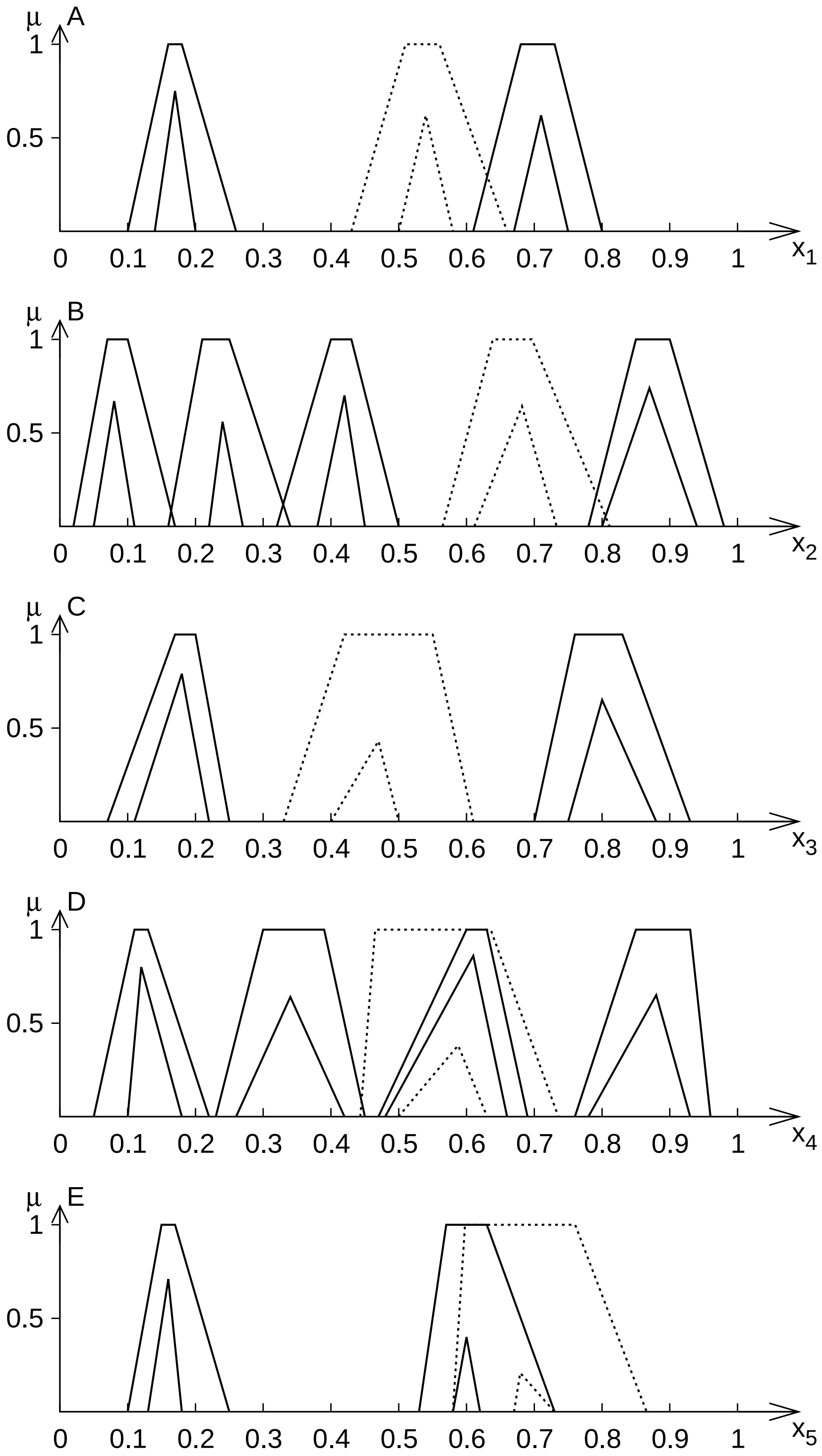


Figure 6.3: Interpolated results from RF interpolation

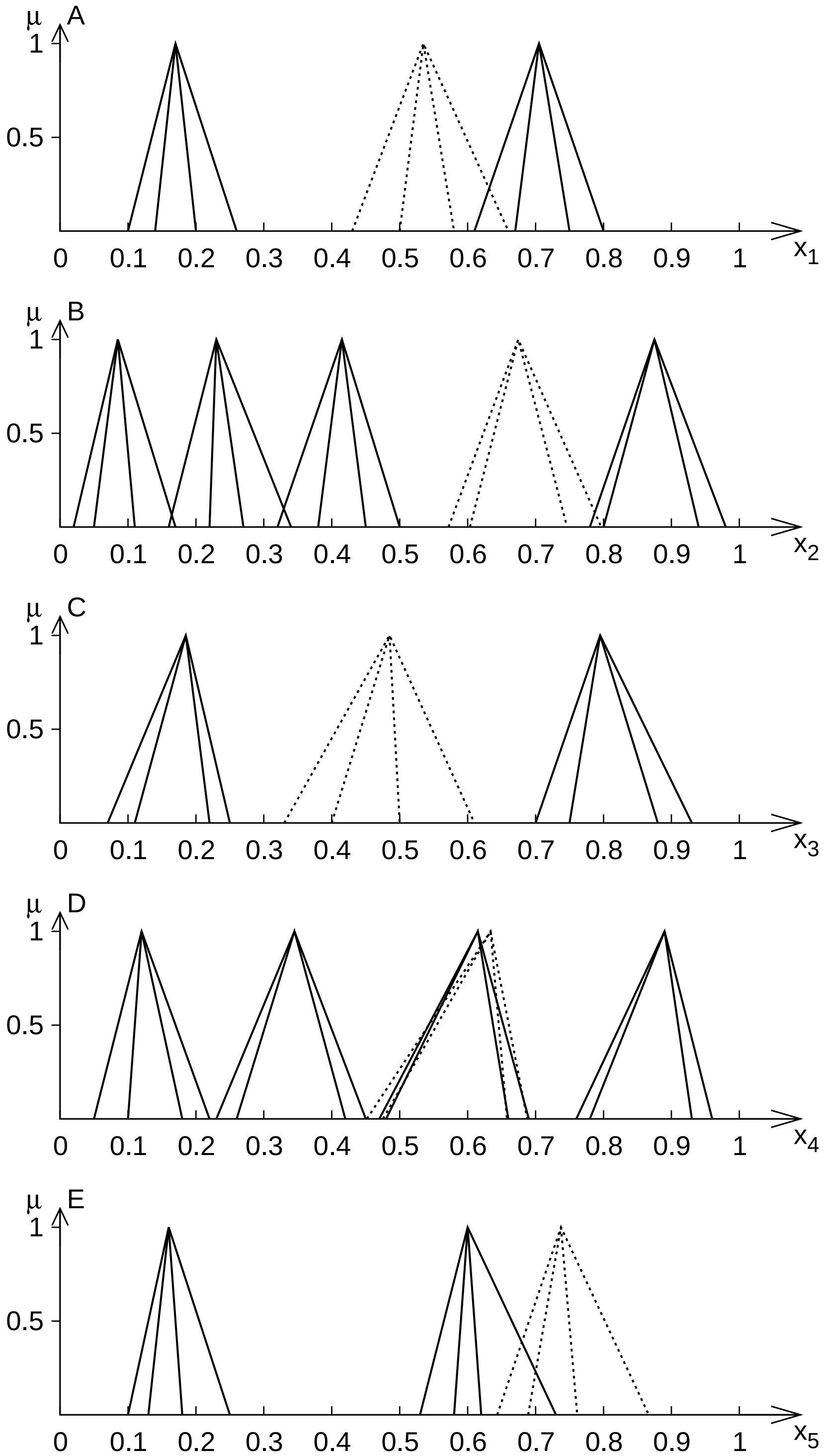


Figure 6.4: Interpolated results from type-2 interpolation

As a consequence, the proposed framework is useful for representing higher order uncertain information, in terms of both data and knowledge, and helps address such types of higher order uncertainty to perform interpolation in a uniformed way.

6.2 Application to Other Realistic Data

In this section, the framework is further evaluated in the application to real datasets for decision making problems. The resultant accuracy reflects the efficacy of this framework.

6.2.1 Decision Making Techniques

Decision making [24, 83, 124] is one of the most important activities for real-world applications of intelligent systems [78]. With given domain knowledge, the task of decision making is to get an optimal or a near optimal solution from input information using an inference procedure. That is, the subject of decision making is the study of how decisions are actually made and how they can be made more successfully [104]. Generally, there are three ways to make a decision in a complex environment [72]:

- by building a mathematical model;
- by seeking human expert advice;
- by building a computational model or an expert system.

Among these, building an accurate mathematical model to describe the complex environment is a good way. However, accurate mathematical models almost always neither exist nor can be derived for all complex environments because the domain may not be thoroughly understood. The first method is therefore limited and when it fails, an alternative for making a good decision is to seek human expert help. However, the cost of querying an expert at any time may be high, and there may be no human experts available when the decision must be made.

Expert systems have been widely used in domains for which the first two methods are not suitable [64, 123]. The knowledge base in an expert system can grow incrementally and can be updated dynamically, so that the performance of an expert system will become better and better as it develops. Also, the expert system approach can integrate expertise from many fields, reduce the cost of query, lower and probability of danger occurring, and provide quicker response [61].

6.2.2 Uncertainty in Decision Process

FS theory is more and more frequently used in expert systems, because of its simplicity and similarity to human reasoning. Most fuzzy expert systems can be seen as special rule-based systems that use fuzzy logic. A fuzzy rule-based expert system contains fuzzy rules in its knowledge base and derives conclusions from the inputs and the fuzzy reasoning process [159, 225]. It usually predefines MFs and fuzzy inference rules to map numerical data onto linguistic variables and to make fuzzy reasoning work, where the linguistic variables are usually defined as FSs with appropriate MFs.

The generation of fuzzy rules from numerical data consists of two phases: the partition of the input spaces into fuzzy subspaces and the determination of the shapes of MFs [80]. This procedure can be achieved by performing a fuzzy partition of the input spaces dividing each universe of discourse into a number of equal or unequal partitions, selecting a type of MF, and assigning one FS to each subspace [43, 95]. For example, these FSs may have linguistic variables such as S: small; MS: medium small; M: medium; ML: medium large and L: large, as shown in Figure 6.5.

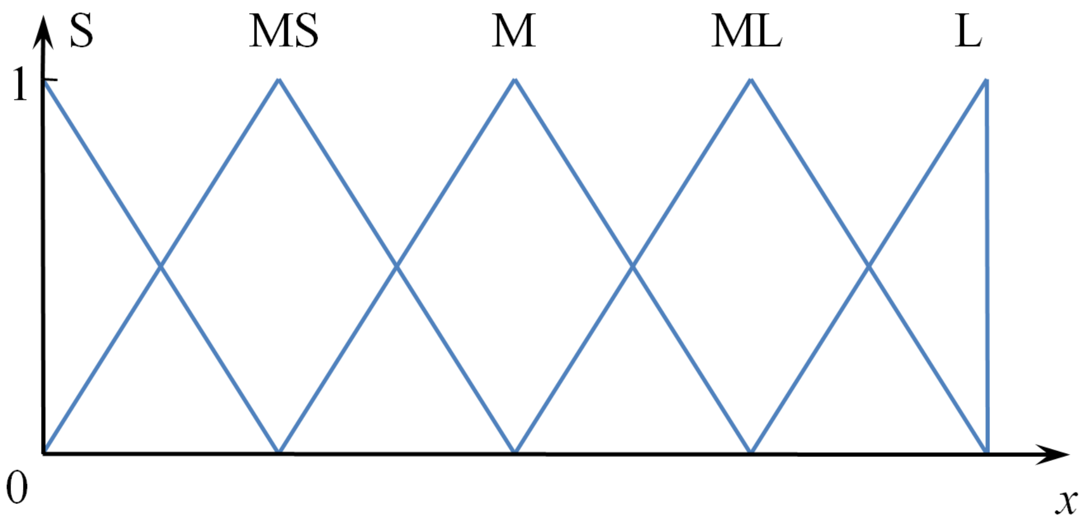


Figure 6.5: Linguistic variables associated with FSs

For a fuzzy partition, an element does not need to be associated with a single region, but has a set of MFs that indicate the extent to which it is regarded as belonging to each of the regions [62]. Usually, the partition can be generated from the advice of human experts. Experts can define a number of FSs for each variable, which are interpreted as linguistic variables and shared by all of the rules

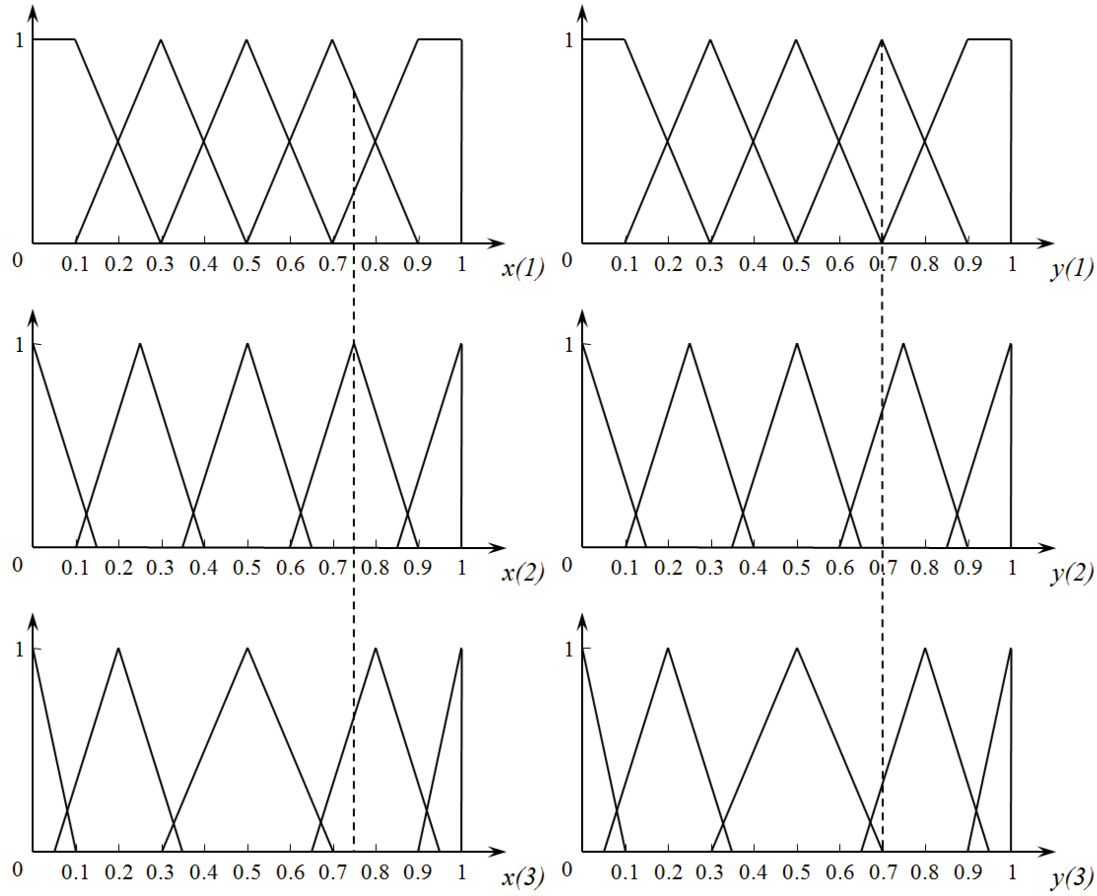
[67]. However, this procedure relies heavily on the opinions of experts, who must have a comprehensive and detailed understanding of the problem at hand. Such opinions are often subjective and/or inconsistent between different individuals. In this case, individuals may have a different understanding for the same information and different experiences in the area of a current problem, so that different kinds of expertise may be obtained from different experts.

Reconsider the popular *tomato problem*: three experts are required to give their opinions on the concepts “colour” and “ripeness”, and their relations, and provide the corresponding fuzzy regions. For simplicity, all of the relevant variable values are normalised into the interval $[0,1]$. Figure 6.6 shows this example where the domain intervals of x and y are divided into five regions, respectively. Each region of x in turn denotes “green”, “green-yellow”, “yellow”, “yellow-red”, and “red”, while each region of y in turn denotes “unripe”, “almost unripe”, “half ripe”, “almost ripe”, and “ripe”. Here, the adopted fuzzy regions for the consequent variable are the same as those of the antecedent variable, and the shape of each MF is triangular. Of course, other divisions of the domain regions and other shapes of MFs are also possible.

Given a pair of input-output data “x is 0.75, y is 0.7”, the degrees of this data pair can be determined by calculating the intersections of each fuzzy region. Since an element can belong to different regions with different degrees, the maximum degree is then chosen to assign a data pair. As a result, one rule can be obtained from one pair of desired input-output data. For the given data pair, a fuzzy rule is therefore generated as

If a tomato is yellow-red, then the tomato is almost ripe.

Note that the same rule can be derived from the partitions that are provided by all three experts, however, this rule may have different meanings for different experts. As shown in Figure 6.7, three similar but different fuzzy regions indicate the uncertain opinions with respect to personal understanding. Obviously, this inconsistency reveals the underlying uncertainty involved in the decision process. As discussed previously, words can mean different things to different people, so that a concept may have an uncertain profile for human opinions. In addition, when a phenomena or an event is too complex or too ill to be expressed, experts would be forced to make unclear judgements. Consequently, the decision process is usually



(a) Antecedent variable "colour"   (b) Consequence variable "ripeness"

Figure 6.6: Divisions of fuzzy regions and the corresponding MFs obtained from three experts

accompanied by imprecision and uncertainty that characterise expert judgements or opinions.

Since different partitions of the same set of elements are usually provided, it is relevant to consider obtaining a single consensus partition which summarises the information contained in the separate partitions. Such a consensus partition provides a way of simplifying this information and obtaining an overall view of the relationships within the set of elements. The reason for doing this is that each partition leads to a single decision result, resulting in difficulty with the consensual decision.

Inspired by this observation, it is beneficial to adopt the proposed higher order framework for modelling the underlying uncertainty. The framework can be applied

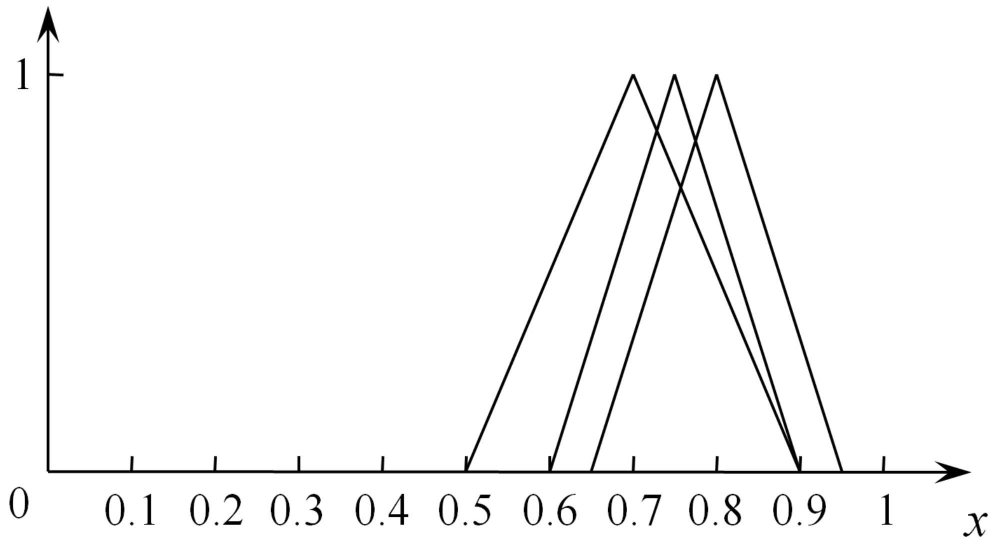


Figure 6.7: Different opinions for the concept "yellow-red"

to construct a consensus representation for characterising a given concept, where the representation can be in the form of RF sets or type-2 FSs. Then, higher order FRI can be implemented to derive the resultant solution. Due to the fact that insufficient training data may result in insufficient rule sets, "gaps" will therefore exist in the generated rule base. As shown in Figure 6.8, when testing data occurs in the middle block, no classical inference methods can derive a result. Therefore, the FRI technique is utilised here.

6.2.3 Performance Evaluation

In order to evaluate the proposed framework, an application to real datasets is provided. In this subsection, a systematic analysis of the performance of the framework for training data and testing data is examined by the UCI *servo* dataset [3], which is made up of 167 instances with four antecedent variables and one consequent variable. The proposed RF sets and RF interpolation are utilised for problem solving here. Further experimental results on other datasets will be presented in Section 6.2.5.

6.2.3.1 Partition of Problem Space

The task of generating and learning fuzzy rules from numerical data is to model the input-output behaviour of a certain system. Without loss of generality, the system

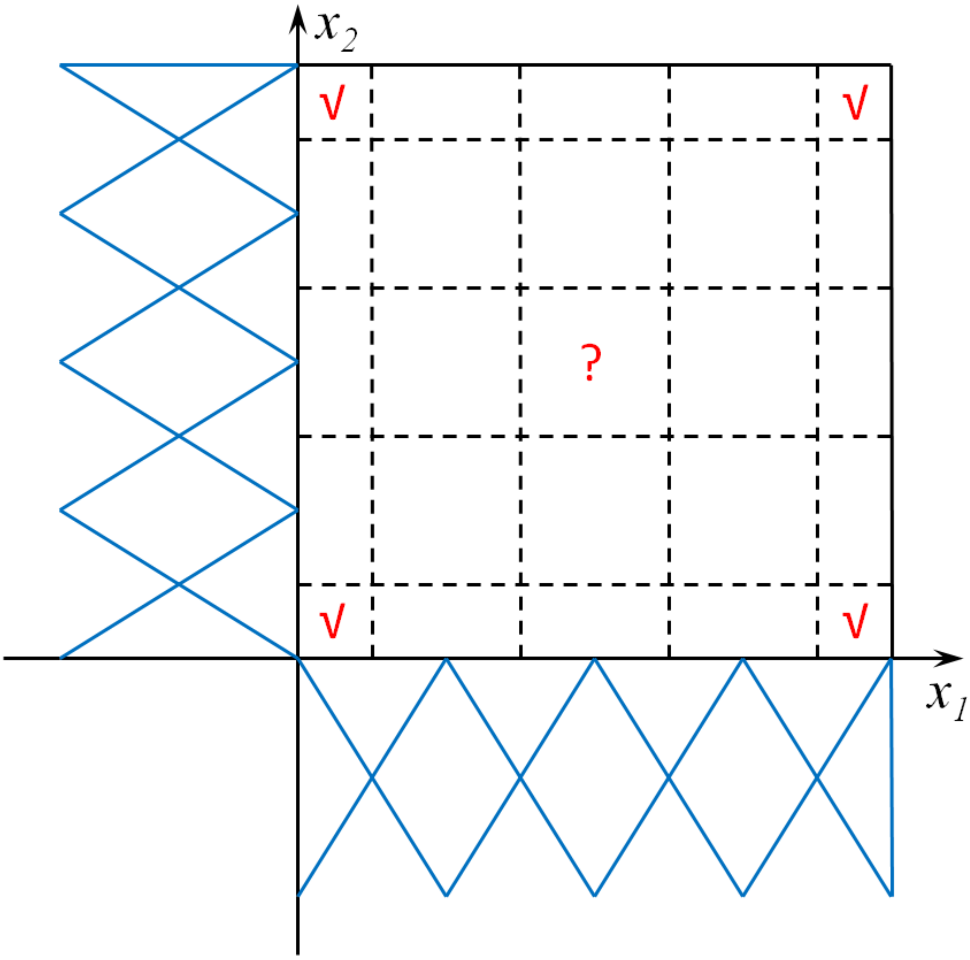


Figure 6.8: An example of the fuzzy partition with "gaps"

to be modelled here is assumed to be a multiple-input-single-output system with M inputs, $\{x_j | j = 1, \cdots, M\}$, and a single output y. A fuzzy rule R_i, $i = 1, \cdots, N$, for such a system is represented as

$$\text{If } x_1 \text{ is } A_{i1}, \cdots, x_j \text{ is } A_{ij}, \cdots, x_M \text{ is } A_{iM}, \text{ then } y \text{ is } B_i.$$

where $x_1, \cdots, x_M$ are the underlying linguistic variables, jointly defining an M-dimensional input space. A_{ij} is the FS of the corresponding antecedent x_j, while B_i is the FS of the consequent y.

In order to generate a set of fuzzy rules, each input/output space is divided into K ($K \geq 2$) subspaces. For simplicity, each variable is normalised into real numbers in the unit interval $[0, 1]$ and divided into K fuzzy regions with the corresponding FSs calculated by the triangular-shaped or trapezoidal-shaped MFs. This is illustrated in

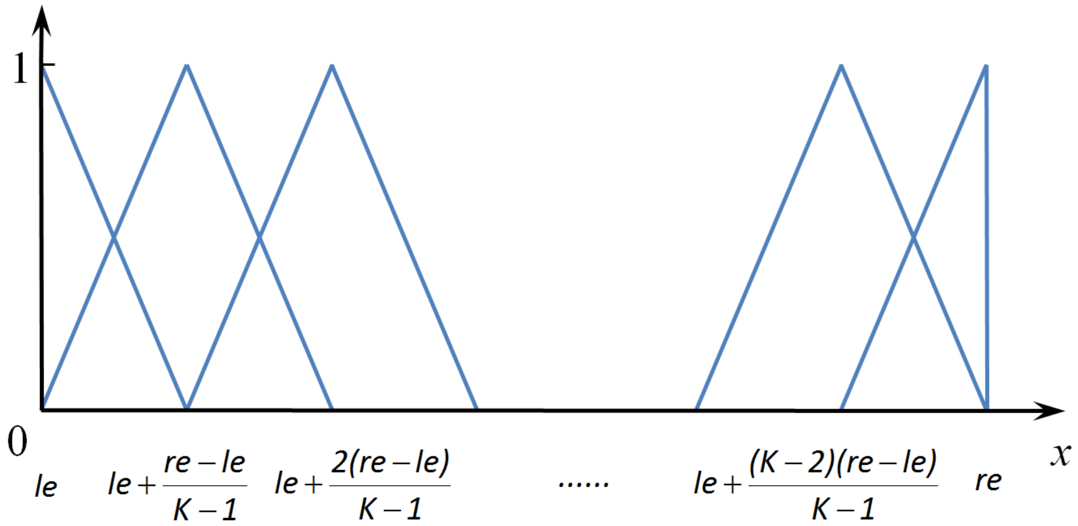


Figure 6.9: Partitioning of each input/output space

Figure 6.9, where le and re denote the left and right extreme values of variable x_j, respectively. The vertex location of a triangle is determined by its position in the K partition. Any membership value of x_j in a new input below le or above re is set to 1.

Suppose that three experts are required to provide their opinions for partitioning the input/output space. Due to the fact that all the antecedent variables are categorical data, the number of partitions is determined by the corresponding number of categories. Thus, the antecedent variables x_1, x_2 and x_4 are divided into five regions, while x_3 is divided into four regions. For simplicity, no uncertainty is assumed to be involved in these variables. That is, the opinions of partitions from all three experts are assumed to be identical. On the other hand, the range of the consequent variable is from 0.13 to 7.10, so uncertainty is considered in this variable. Three kinds of expertise are therefore chosen for representing different partitions, which are decided by

$$\begin{cases} le = 0, \ re = 1 \\ le = 0.05, \ re = 0.95 \\ le = 0.1, \ re = 0.9 \end{cases} \tag{6.1}$$

resulting in three types of fuzzy regions, as shown in Figure 6.10. These partitions are then used to determine a set of fuzzy rules that model the relationships between the input-output data pairs.

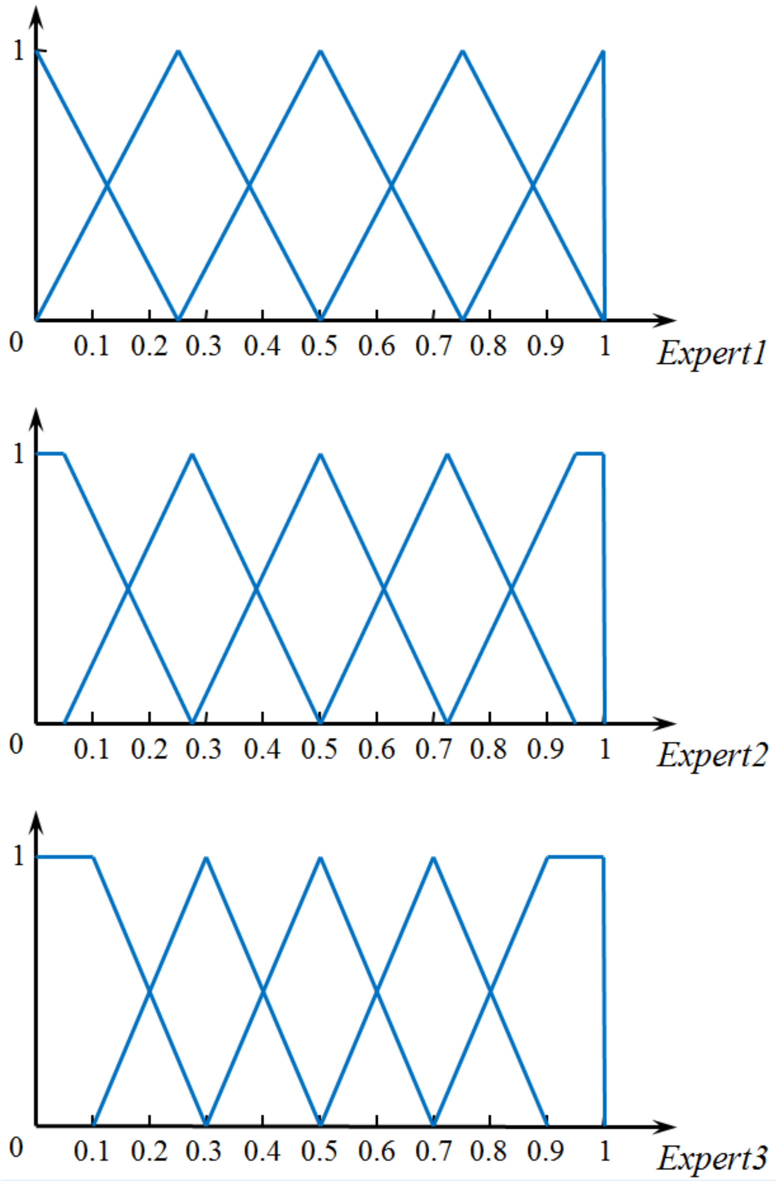


Figure 6.10: Fuzzy regions provided by three experts

Note that only the consequent is regarded as involving uncertainty, this is for the purpose of computational simplicity in the experimental illustration. However, once the uncertainty is included in a certain system, the proposed framework can be applied for representing the underlying uncertainty and deriving a consensual outcome. In this case, the efficacy of the framework is evaluated by comparing the obtainable result with those from three human experts. In order to reflect the gradualness in the improvement of fuzzy partition quality, six different fuzzy partitions are tested where the consequent is each divided into K ($K = 2, \ldots, 7$) fuzzy subsets, where $K = 2$ represents a very rough partition, while $K = 7$ represents a very detailed partition for the servo dataset, which contains only 167 instances. The purpose of investigating the use of different dataset partitions is to examine the performance of the proposed framework for fine fuzzy partitions as well as coarse fuzzy partitions.

6.2.3.2 Generation of Fuzzy Rules

Given data pairs, a fuzzy rule base can be formed by creating a rule that best covers a certain input-output data pair [179]. This work uses this method owing to its technical maturity and conceptual simplicity, where the region with maximum degree is assigned to each input-output data pair.

As a result, three rule bases can be constructed from the opinions of each expert. Next, an RF rule base can be built on top of these rule bases. That is, each fuzzy region triple is aggregated into an RF set, where the uncertainty is described by the lower and upper approximations. Based on the membership of LA, a given datum is then allocated to the region with maximum membership degree. Since the LA characterises the grade of certainty, a higher degree indicates a higher certainty. Such a rule base includes the first and higher order types of uncertainty and represents them as RF sets. Also, they are considered in the process of interpolation in order to obtain better inference conclusions.

Note that the rules generated in this way are logical conjunctive rules, i.e., rules in which the conditions of the IF part must be met simultaneously in order for the result of the THEN part to occur. For the problem considered here, only conjunctive rules are required since the antecedents are different components of a single input vector.

6.2.3.3 Implementation of Interpolation

The original T-FRI approach is adopted for interpolation in the single expert rule base, while the proposed RF interpolation is used for the RF rule base. This process is implemented repeatedly for each partition.

In addition, N ($N = 2, \ldots, 6$) fuzzy rules are chosen as the closest rules to respectively interpolate the conclusions. The purpose of using different numbers of closest rules is to analyse their influence on the construction of the outcome.

6.2.3.4 Evaluation of Accuracy

In order to make a comprehensive comparison, each resultant set is defuzzified to a crisp value using its Rep. Root-mean-square error (RMSE) is adopted to calculate the accuracy:

$$\epsilon_{RMSE} = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(O_k - G_k)^2} \tag{6.2}$$

where O_k and G_k denote the kth testing output value and its corresponding ground truth (the consequent of the testing data), respectively.

Ten times 10-fold cross-validation [109, 162] is then employed to evaluate the generalisation ability of the proposed approach. The servo dataset with 167 instances is randomly divided into 10 subsets for similar size, where three of which contain 16 instances and seven of which contain 17 instances. One single subset is maintained as the validation data for testing, while the remaining 9 subsets are used for training. This is then repeated 10 times with each of the subsets used as the testing data, and the rest as the training data. The 10 results from the folds are averaged to produce a single accuracy value. The process is then repeated 10 times by initialising different, randomly assigned initial 10 subsets.

6.2.4 Discussion of Results

Tables 6.2, 6.3 and 6.4 list the results of the averaged RMSE in terms of 10 times 10-fold cross-validation in relation to K partitions and N closest rules. Paired t-test results with significance level of 0.05 are also identified in these tables with the achieved accuracies of the RF approach as reference, those significantly better, worse and no difference are marked with “(v)”, “(*)” and “(-)”, respectively.

As reflected in these tables, the accuracies from three separate expert rule bases are unstable. That is, the opinions from an expert can perform well in certain partitions, and badly in others. Theoretically, this is acceptable as someone is only an expert in a particular field, namely, the necessary expertise may be only available for a certain concept. However, this leads to difficulty for making decisions in practical applications. Since different opinions result in better or worse accuracy, it is difficult to conclude which one should be chosen.

However, it is obvious that the accuracy obtained by the proposed RF approach is generally higher than the use of T-FRI directly over the three single-expert rule

Table 6.2: Accuracies (RMSE×100) of the servo data for $K = 2$ and $K = 3$

	Partition = 2				Partition = 3			
Closest rules	Expert 1	Expert 2	Expert 3	RF	Expert 1	Expert 2	Expert 3	RF
2	25.12(*)	24.51(*)	23.90(*)	22.72	13.03(v)	13.60(-)	13.91(*)	13.54
3	25.96(*)	25.37(*)	24.79(*)	23.65	14.53(-)	14.97(*)	15.27(*)	14.78
4	26.46(*)	25.88(*)	25.31(*)	24.18	15.69(-)	16.04(*)	16.32(*)	15.85
5	26.25(*)	25.65(*)	25.06(*)	23.90	15.23(-)	15.57(*)	15.87(*)	15.36
6	26.20(*)	25.60(*)	25.00(*)	23.83	15.13(-)	15.45(*)	15.74(*)	15.29

Table 6.3: Accuracies (RMSE×100) of the servo data for $K = 4$ and $K = 5$

	Partition = 4				Partition = 5			
Closest rules	Expert 1	Expert 2	Expert 3	RF	Expert 1	Expert 2	Expert 3	RF
2	11.46(*)	11.10(*)	11.45(*)	10.55	10.66(*)	10.31(*)	10.42(*)	9.93
3	12.79(*)	12.79(*)	13.10(*)	12.29	11.99(*)	11.91(*)	12.21(*)	11.57
4	13.42(*)	13.67(*)	14.03(*)	13.15	12.65(*)	12.78(*)	13.27(*)	12.47
5	12.78(*)	13.00(*)	13.40(*)	12.42	11.81(*)	12.00(*)	12.52(*)	11.64
6	12.59(*)	12.78(*)	13.22(*)	12.19	11.60(*)	11.78(*)	12.29(*)	11.36

Table 6.4: Accuracies (RMSE×100) of the servo data for $K = 6$ and $K = 7$

	Partition = 6				Partition = 7			
Closest rules	Expert 1	Expert 2	Expert 3	RF	Expert 1	Expert 2	Expert 3	RF
2	10.20(*)	9.95(*)	10.34(*)	9.86	9.67(-)	9.93(*)	10.11(*)	9.80
3	11.70(*)	11.63(*)	12.01(*)	11.44	11.37(-)	11.52(*)	11.75(*)	11.30
4	12.46(*)	12.51(*)	13.00(*)	12.29	12.34(*)	12.36(*)	12.60(*)	12.09
5	11.69(*)	11.76(*)	12.30(*)	11.52	11.50(*)	11.52(*)	11.82(*)	11.24
6	11.50(*)	11.50(*)	12.07(*)	11.24	11.29(*)	11.30(*)	11.59(*)	11.01

bases. This reflects an important advantage of the proposed framework in that more considered information produces better results. The uncertainty provides useful information on depicting a concept. Instead of discarding the uncertainty, it is better to capture and represent such uncertainty.

In addition, the effect of the numbers of K and N can also be seen in the tables. First, when $K = 2$, the accuracies from all four rule bases are very low. This shows that a rough partition will lead to difficulty in identifying the difference between similar training data, resulting in poor accuracy. As K increases, more detailed partitions are generated for better differentiating data pairs. This is verified by the results in that the errors fall as the value of K is increased, as shown in Figure 6.11.

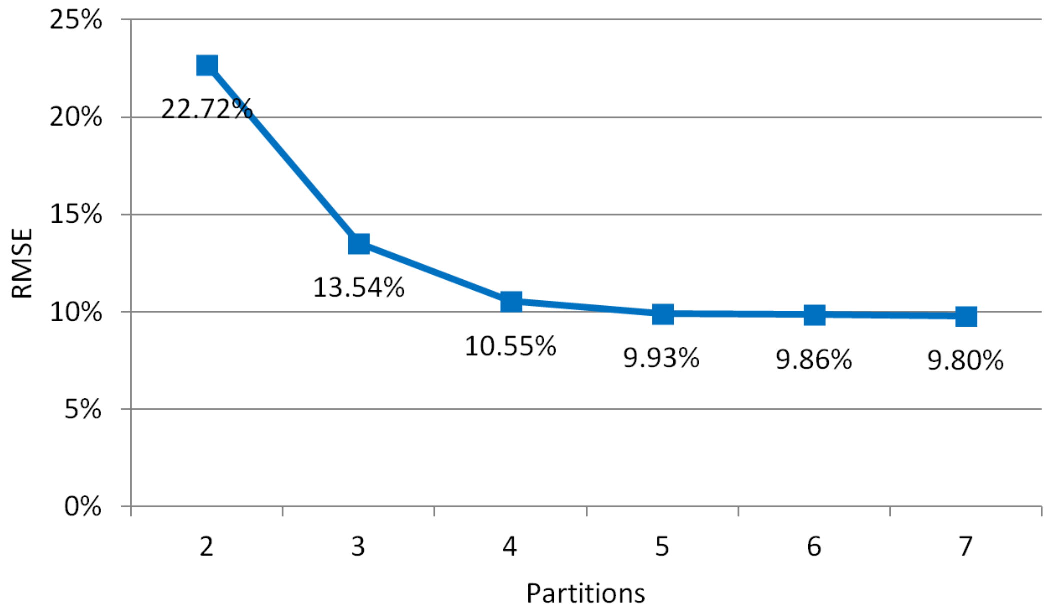


Figure 6.11: Accuracies of RF approach for different partitions when $S = 2$

In contrast to the trend for the increase of K, the change of the value for N only slightly affects the accuracy. As noted previously, interpolation requires at least two closest neighbouring rules. Here, the "closest" ensures that a given observation is as close to the antecedents of the neighbouring rules as possible, as well as the interpolated result to the consequents of those rules. Naturally, for more than two rules, results are expected to be better. However, proximity is measured by the averaged distance for all the antecedent variables. That is to say, the chosen rules are not the "closest" ones for every antecedent variable. This causes a variation for the resultant accuracies, as shown in Figure 6.12.

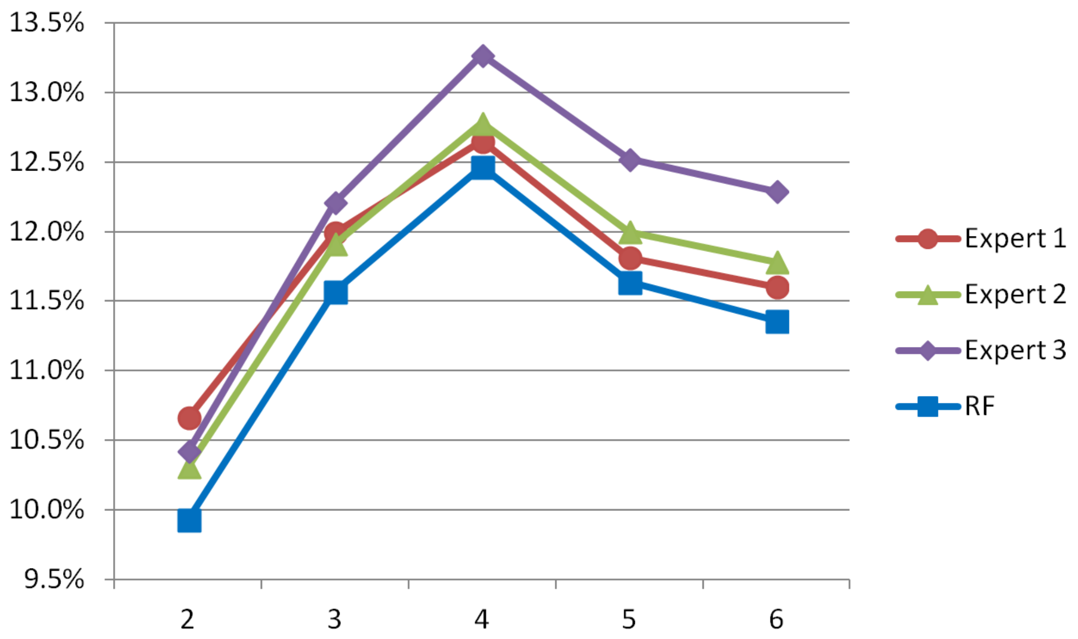


Figure 6.12: Accuracies for different numbers of selected rules when $K = 5$

6.2.5 Further Applications with More Datasets

The work is also evaluated on the *Yacht Hydrodynamics* dataset, *Airfoil Self-Noise* dataset, *Concrete Compressive Strength* dataset, and *Housing* dataset [3]. In general, the uncertainty is considered in every attribute. The details of the calculation are omitted here. The obtainable results are listed in Tables 6.5 - 6.16, respectively. Generally, the accuracies from the RF approach are better or at least comparable to the single-expert ones. These results also reflect the effectiveness of the work.

Consequently, according to the experiment results, the proposed framework can not only help to represent the uncertainty in knowledge, but can also assure the decision accuracy by exploiting the uncertain information.

6.2.6 Further Applications with More Experts

This work is further evaluated with more experts generated using a flexible approach than that proposed in Figures 6.9 and 6.10. In particular, Expert 1 is defined as a generator using Equation (6.1) (i.e., $le = 0$, $re = 1$), while the others are not predefined. Instead, they are randomly constructed from Expert 1 (i.e., $le \in (0, 1)$, $re = 1 - le \in (0, 1)$), resulting in four other different fuzzy regions. The obtainable

Table 6.5: Accuracies (RMSE×100) of the yacht data for $K = 2$ and $K = 3$

	Partition = 2				Partition = 3			
Closest rules	Expert 1	Expert 2	Expert 3	RF	Expert 1	Expert 2	Expert 3	RF
2	31.37(*)	30.99(*)	30.63(*)	27.64	20.54(v)	21.02(v)	21.40(-)	21.42
3	31.15(*)	30.74(*)	30.35(*)	26.83	19.65(v)	20.14(-)	20.60(*)	20.11
4	31.10(*)	30.68(*)	30.26(*)	26.43	19.30(-)	19.77(-)	20.14(*)	19.62
5	31.00(*)	30.56(*)	30.13(*)	26.04	19.31(-)	19.76(*)	19.84(*)	19.23
6	30.90(*)	30.46(*)	30.02(*)	25.80	19.50(*)	19.93(*)	19.64(*)	19.10

Table 6.6: Accuracies (RMSE×100) of the yacht data for $K = 4$ and $K = 5$

	Partition = 4				Partition = 5			
Closest rules	Expert 1	Expert 2	Expert 3	RF	Expert 1	Expert 2	Expert 3	RF
2	15.95(*)	16.66(*)	20.69(*)	15.56	14.75(v)	16.76(-)	17.13(-)	17.05
3	16.28(*)	16.88(*)	20.66(*)	15.41	16.39(v)	18.04(*)	18.26(*)	17.51
4	17.51(*)	18.52(*)	20.97(*)	16.31	17.04(v)	18.99(*)	19.23(*)	17.53
5	18.46(*)	19.64(*)	21.31(*)	17.18	16.58(v)	19.02(*)	19.34(*)	16.87
6	18.72(*)	19.82(*)	21.20(*)	17.31	16.53(-)	18.21(*)	18.52(*)	16.59

Table 6.7: Accuracies (RMSE×100) of the yacht data for $K = 6$ and $K = 7$

	Partition = 6				Partition = 7			
Closest rules	Expert 1	Expert 2	Expert 3	RF	Expert 1	Expert 2	Expert 3	RF
2	14.98(-)	14.98(-)	17.90(*)	14.87	14.92(-)	15.15(*)	14.87(-)	14.86
3	15.47(v)	16.71(*)	18.30(*)	16.46	15.55(*)	15.65(*)	14.85(v)	15.25
4	14.80(v)	17.34(*)	18.46(*)	17.13	15.02(*)	15.34(*)	15.76(*)	14.88
5	14.72(v)	16.40(-)	18.00(*)	16.42	15.05(-)	15.75(*)	16.15(*)	15.14
6	14.72(v)	16.10(-)	17.87(*)	16.10	15.23(v)	16.35(*)	16.01(*)	15.52

Table 6.8: Accuracies (RMSE×100) of the airfoil data for $K = 2$ and $K = 3$

	Partition = 2				Partition = 3			
Closest rules	Expert 1	Expert 2	Expert 3	RF	Expert 1	Expert 2	Expert 3	RF
2	19.31(*)	19.53(*)	19.76(*)	18.94	17.59(-)	17.62(-)	17.56(-)	17.55
3	18.73(*)	18.88(*)	19.05(*)	18.49	16.91(*)	16.70(-)	16.50(v)	16.66
4	18.09(*)	18.17(*)	18.26(*)	17.98	16.32(*)	16.08(-)	16.05(-)	16.07
5	17.82(*)	17.86(*)	17.91(*)	17.77	16.31(*)	15.81(-)	15.74(-)	15.73
6	17.69(*)	17.73(*)	17.78(*)	17.66	16.10(*)	15.60(-)	15.51(-)	15.53

Table 6.9: Accuracies (RMSE×100) of the airfoil data for $K = 4$ and $K = 5$

Closest rules	Partition = 4				Partition = 5			
	Expert 1	Expert 2	Expert 3	RF	Expert 1	Expert 2	Expert 3	RF
2	15.86(v)	16.30(*)	15.97(v)	16.19	14.03(v)	14.85(*)	15.66(*)	14.64
3	14.80(v)	15.48(-)	15.04(v)	15.47	13.02(v)	13.98(*)	14.79(*)	13.85
4	14.30(v)	15.01(-)	14.68(v)	14.98	12.66(v)	13.57(*)	14.35(*)	13.49
5	13.98(v)	14.76(-)	14.32(v)	14.69	12.48(v)	13.40(*)	14.08(*)	13.32
6	13.81(v)	14.50(-)	14.05(v)	14.43	12.40(v)	13.36(*)	13.85(*)	13.27

Table 6.10: Accuracies (RMSE×100) of the airfoil data for $K = 6$ and $K = 7$

	Partition = 6				Partition = 7			
Closest rules	Expert 1	Expert 2	Expert 3	RF	Expert 1	Expert 2	Expert 3	RF
2	13.20(v)	13.66(*)	14.56(*)	13.48	12.25(v)	13.53(*)	14.42(*)	13.11
3	12.32(v)	12.77(*)	13.66(*)	12.60	11.45(v)	12.75(*)	13.62(*)	12.13
4	12.02(v)	12.49(*)	13.34(*)	12.29	11.17(v)	12.48(*)	13.28(*)	11.68
5	11.88(v)	12.41(*)	13.14(*)	12.10	11.11(v)	12.34(*)	13.11(*)	11.45
6	11.86(v)	12.35(*)	13.10(*)	11.97	11.16(v)	12.32(*)	13.05(*)	11.43

Table 6.11: Accuracies (RMSE×100) of the concrete data for $K = 2$ and $K = 3$

Closest rules	Partition = 2				Partition = 3			
	Expert 1	Expert 2	Expert 3	RF	Expert 1	Expert 2	Expert 3	RF
2	19.50(*)	19.56(*)	19.70(*)	19.42	18.06(v)	18.72(*)	19.04(*)	18.32
3	18.86(*)	18.96(*)	19.04(*)	18.86	17.38(v)	17.85(*)	18.02(*)	17.49
4	18.49(*)	18.47(*)	18.50(*)	18.41	17.15(v)	17.38(*)	17.47(*)	17.26
5	18.20(*)	18.18(*)	18.15(-)	18.14	17.03(-)	17.05(-)	17.25(*)	17.07
6	18.10(*)	18.06(-)	18.03(-)	18.05	16.82(-)	16.94(*)	17.01(*)	16.85

Table 6.12: Accuracies (RMSE×100) of the concrete data for $K = 4$ and $K = 5$

	Partition = 4				Partition = 5			
Closest rules	Expert 1	Expert 2	Expert 3	RF	Expert 1	Expert 2	Expert 3	RF
2	16.66(*)	16.49(*)	16.16(-)	16.06	14.61(v)	16.13(*)	16.15(*)	15.84
3	15.87(*)	15.73(*)	15.44(-)	15.52	14.10(v)	15.41(*)	15.41(*)	15.21
4	15.41(*)	15.31(*)	15.00(v)	15.21	13.86(v)	14.94(*)	15.00(*)	14.79
5	15.21(*)	15.03(*)	14.75(v)	14.96	13.67(v)	14.75(-)	14.84(*)	14.69
6	15.09(*)	14.84(-)	14.62(v)	14.85	13.54(v)	14.65(-)	14.62(-)	14.63

Table 6.13: Accuracies (RMSE×100) of the concrete data for $K = 6$ and $K = 7$

	Partition = 6				Partition = 7			
Closest rules	Expert 1	Expert 2	Expert 3	RF	Expert 1	Expert 2	Expert 3	RF
2	14.89(*)	14.17(v)	16.06(*)	14.65	14.57(v)	14.87(*)	15.95(*)	14.69
3	14.28(*)	13.73(v)	15.50(*)	14.01	14.05(v)	14.43(*)	15.35(*)	14.21
4	14.03(*)	13.65(v)	15.08(*)	13.86	13.62(v)	14.06(*)	14.93(*)	13.82
5	13.65(*)	13.43(-)	14.83(*)	13.45	13.30(v)	13.65(*)	14.63(*)	13.49
6	13.47(*)	13.34(-)	14.66(*)	13.36	13.14(v)	13.49(*)	14.42(*)	13.28

Table 6.14: Accuracies (RMSE×100) of the housing data for $K = 2$ and $K = 3$

	Partition = 2				Partition = 3			
Closest rules	Expert 1	Expert 2	Expert 3	RF	Expert 1	Expert 2	Expert 3	RF
2	18.15(*)	18.25(*)	18.23(*)	17.98	16.93(*)	16.60(-)	16.39(-)	16.45
3	17.83(*)	17.91(*)	17.82(*)	17.74	16.66(*)	16.33(-)	15.96(-)	16.08
4	17.47(*)	17.47(*)	17.39(-)	17.38	16.71(*)	16.21(*)	15.70(-)	15.86
5	17.33(-)	17.31(-)	17.22(-)	17.27	16.67(*)	16.13(*)	15.64(-)	15.71
6	17.26(-)	17.22(-)	17.13(v)	17.21	16.65(*)	16.10(*)	15.54(-)	15.75

Table 6.15: Accuracies (RMSE×100) of the housing data for $K = 4$ and $K = 5$

	Partition = 4				Partition = 5			
Closest rules	Expert 1	Expert 2	Expert 3	RF	Expert 1	Expert 2	Expert 3	RF
2	16.06(*)	15.31(v)	14.56(v)	15.56	14.51(-)	14.34(-)	13.95(v)	14.42
3	15.46(*)	14.89(-)	14.32(v)	14.94	14.00(-)	13.89(-)	13.49(v)	13.97
4	15.32(*)	14.74(-)	14.37(v)	14.78	13.95(-)	13.74(-)	13.57(v)	13.83
5	15.19(*)	14.65(v)	14.39(v)	14.72	14.04(*)	13.82(-)	13.62(v)	13.82
6	15.15(*)	14.70(-)	14.44(v)	14.73	14.09(*)	13.89(-)	13.61(v)	13.86

Table 6.16: Accuracies (RMSE×100) of the housing data for $K = 6$ and $K = 7$

	Partition = 6				Partition = 7			
Closest rules	Expert 1	Expert 2	Expert 3	RF	Expert 1	Expert 2	Expert 3	RF
2	13.71(*)	13.28(-)	13.89(*)	13.33	12.85(v)	13.04(v)	13.75(*)	13.40
3	13.32(*)	13.19(-)	13.55(*)	13.13	12.76(v)	12.90(v)	13.22(-)	13.18
4	13.32(-)	13.20(-)	13.49(*)	13.20	12.76(v)	12.91(-)	13.28(*)	12.95
5	13.43(-)	13.42(-)	13.58(*)	13.41	12.88(v)	13.03(-)	13.32(*)	13.06
6	13.51(-)	13.57(-)	13.66(-)	13.56	12.98(v)	13.19(-)	13.42(*)	13.16

results, which are calculated from the *Servo* dataset, *Yacht Hydrodynamics* dataset, and *Housing* dataset [3], are listed in Tables 6.17 - 6.19, respectively.

Table 6.17: Accuracies (RMSE×100) of the servo data

	Partition = 2					
Closest rules	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	RF
2	20.75(*)	20.41(*)	20.10(*)	20.58(*)	20.25(*)	19.88
4	22.38(*)	22.07(*)	21.78(*)	22.22(*)	21.92(*)	21.22
6	22.08(*)	21.75(*)	21.44(*)	21.92(*)	21.59(*)	20.75
	Partition = 4					
Closest rules	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	RF
2	10.72(*)	10.50(-)	10.71(*)	10.70(*)	10.65(-)	10.54
4	12.67(-)	12.95(*)	13.15(*)	12.82(*)	13.04(*)	12.71
6	11.88(-)	12.10(*)	12.35(*)	12.02(*)	12.25(*)	11.91
	Partition = 6					
Closest rules	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	RF
2	10.00(*)	9.90(*)	10.09(*)	10.12(*)	9.92(*)	9.74
4	11.86(-)	11.98(-)	12.37(*)	11.89(-)	12.08(*)	12.00
6	10.94(v)	11.05(-)	11.52(*)	10.96(v)	11.22(*)	11.14

Again, the accuracies from the RF approach are generally better or at least comparable to the other five single-expert ones. These results also reflect the effectiveness of this work.

6.3 Summary

For evaluation purposes, the proposed framework has been applied to a realistic problem of predicting diarrhoeal disease rates in roadless villages in this chapter. This problem presents itself as a suitable testbed due to its nature of lacking detailed information and comprehensive knowledge. Although different kinds of expertise generally lead to difficulty in determining the final inference outcome, application of the higher order framework consistently results in good performance. This demonstrates the potential of this work in improving the effectiveness of FRI.

To further evaluate the present work, application to other datasets have also been provided in this chapter. A rule-based fuzzy system works on the generation of fuzzy rules from numerical data. However, different expert opinions on fuzzy partitions may result in uncertainty in the overall domain knowledge. The proposed framework provides a good solution by including the uncertainty into the inference process. Experimental results have shown that the exploitation of uncertain knowledge across multiple opinions offered by different experts generates better results than the use of just the expertise offered by a single expert.

Table 6.18: Accuracies (RMSE×100) of the yacht data

	Partition = 2					
Closest rules	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	RF
2	32.07(*)	31.74(*)	31.42(*)	31.91(*)	31.58(*)	30.83
4	31.31(*)	30.89(*)	30.49(*)	31.10(*)	30.69(*)	29.70
6	31.10(*)	30.66(*)	30.23(*)	30.88(*)	30.45(*)	29.40
	Partition = 4					
Closest rules	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	RF
2	15.46(v)	16.03(*)	19.75(*)	15.79(*)	16.05(*)	15.67
4	17.05(*)	17.80(*)	20.32(*)	16.79(*)	17.80(*)	16.41
6	18.13(*)	19.22(*)	20.56(*)	17.68(*)	19.22(*)	17.38
	Partition = 6					
Closest rules	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	RF
2	14.49(v)	14.71(-)	17.49(*)	14.52(v)	16.99(*)	14.85
4	16.50(v)	17.09(-)	18.04(*)	17.15(-)	17.60(*)	17.09
6	15.42(-)	15.79(*)	17.46(*)	15.51(*)	16.24(*)	15.23

Table 6.19: Accuracies (RMSE×100) of the housing data

	Partition = 2					
Closest rules	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	RF
2	18.68(*)	18.83(*)	18.76(*)	18.69(*)	18.73(*)	18.60
4	17.72(*)	17.71(*)	17.59(*)	17.68(*)	17.61(*)	17.49
6	17.62(*)	17.59(*)	17.44(*)	17.57(*)	17.47(*)	17.34
	Partition = 4					
Closest rules	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	RF
2	15.80(*)	15.00(-)	14.22(v)	15.50(*)	14.37(v)	15.14
4	14.96(*)	14.37(*)	13.91(v)	14.68(*)	14.14(-)	14.25
6	14.87(*)	14.30(-)	13.97(v)	14.58(*)	14.11(-)	14.22
	Partition = 6					
Closest rules	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	RF
2	13.38(*)	12.99(-)	13.51(*)	13.32(*)	13.43(*)	12.95
4	12.88(-)	12.77(-)	13.00(*)	12.91(-)	12.96(*)	12.82
6	13.06(-)	13.09(*)	13.12(*)	13.19(*)	13.16(*)	12.98

Chapter 7

Theoretical Extension

THE goal in group decision making (GDM) is to ensure that the best decision is made with respect to the available information and knowledge possessed by all group members. However, different types of uncertainty may influence both the assessment of the individual views and the derivation of the overall group-level solution. The difficulty in such decision-making may escalate if the views of all individuals only cover part of the problem space. Systems capable of reasoning through fuzzy interpolation can help in this, as argued previously.

This chapter presents an extended approach for achieving GDM via fuzzy interpolation. Individual preferences are firstly aggregated by means of a method learned on RF set theory, and RF interpolation is then applied to derive the group-level conclusion. Experimental investigations are carried out and the results are presented to demonstrate the efficacy of the proposed work in guaranteeing the overall decision accuracy.

7.1 Group Decision Making Problem

GDM [69, 97] is a process where a number of individuals attempt to reach a consensus on a certain decision. A group solution is the one that is the most acceptable to all the individuals concerned as a whole. In GDM, both the individuals in the group and the group at large jointly make decisions. To do this, individuals need to express their judgements among a set of alternative opinions. Different types of uncertainty may however, influence both the assessment of the individual views and the derivation of the overall group-level solution [127]. These include the following factors: (1) An individual's role (weight) in the generation of the group solutions, since there may be a group leader or leaders who play more important roles in a particular GDM process. (2) An individual's preference for possible decision alternatives, since individuals may have a different understanding for the same information and different experiences in the area of current decision problems. (3) An individual's use of criteria for assessing alternatives, since individuals may often have different judgements in comparing the importance between those criteria. All such types of uncertainty translate into difficulties in determining the final solution by the group. In addition, there are many situations where the potential decision alternatives may be ordered and even depicted on an underlying continuum [98]. Each individual may have an optimum or most preferred position on the continuum. Obviously, the closer any given alternative lies to the optimum, the more it may be preferred over another. Sometimes, an individual's optimum may be located between two distinct alternatives. That is, a different preference may appear beyond given alternatives, leading to the difficulty of making a consensual decision.

It is well-known that human judgement including preferences is often subjective, vague and imprecise. Fuzzy systems play an important role in decision making and offer a flexible framework for GDM. Indeed, fuzzy rules are often employed by human beings to make decisions. Such rules use a series of IF-THEN statements to describe what action should be taken in terms of the currently observed information. They are widely used in FISs to perform decision making according to given individuals' preferences.

The compositional rule of inference [214] offers an effective mechanism to deal with fuzzy inference for dense rule bases. Given such a rule base and an observation that is at least partially covered by the rule base, the conclusion can be inferred from certain rules that intersect with the observation. However, for the case where a fuzzy rule base contains "gaps" (termed: sparse rule base [183]), if a given observation has no overlap with the antecedent values of any rule, conventional fuzzy inference methods cannot derive a conclusion. This is of particular significance when a given preference lies between two known alternatives in GDM. Fortunately, using FRI, certain decisions may still be reached. However, different types of uncertainty may influence both the assessment of the individual views and the derivation of the overall group-level solution in GDM. To cope with such uncertain information and knowledge, the proposed higher order fuzzy representation may be helpful.

An FRI technique for GDM is herein proposed in order to better address the underlying relative uncertainty, thereby determining appropriate decisions. For each criterion, the OWA operator [191] is employed to decide each individual's role. Then, aggregation of individuals' preferences is performed by means of the proposed RF set theoretic approach. Finally, the RF interpolation is utilised to enable required interpolative reasoning.

7.2 Extended Rough-Fuzzy Set Representation

The objective of *aggregation* is to combine individuals' preferences into an overall aggregated value so that the final decision takes into account all individuals' contributions. Different but similar opinions are usually aggregated to provide more robust solutions. The particular concern of this work is to deal with the situations where conclusions cannot be inferred but may be interpolated when given uncertain observations have no overlap with any rules.

One possible approach is to interpolate all the conclusions separately with respect to each given observation first and then, to derive the final solution by aggregating all the individual conclusions. This approach is hereafter denoted as the IA method, standing for interpolation before aggregation. However, as outlined previously, the first step of interpolation requires the computation of the closest rules from a given rule base. A distance measure needs to be calculated in order to estimate the proximity between each rule antecedent and observation antecedent. This implies a time complexity of $O(xmn)$, where x is the number of observations to be interpolated, m is the number of antecedent variables, and n is the number of fuzzy rules involved in a rule base. An alternative approach creates an artificial observation by aggregating all the observations first and then, to derive the final solution by performing interpolation over this artificial observation. For obvious reasons, this approach is hereafter denoted as the AI method, which has an overall time complexity of $O(mn)$. The reduction in computation complexity is significant, especially when the number of observations becomes large. Consequently, the AI method is employed herein for problem solving while the results are compared to those obtainable by the IA method. The following presents the theoretical extension of the RF sets that used for AI approach.

In dealing with individuals' preferences, the pessimistic means is to aggregate such preferences by an intersection operation, in order to ensure that all preferences are satisfied. Opposite to this, the optimistic means is to create the artificial overall preference by performing a union operation in an effort to satisfy at least a single preference. To enable the representation of different types of uncertainty, RF sets can be used to support the aggregation. Thus, Definition 4.2 can be applied for situations where all opinions share a common point. Unfortunately, for many instances, individuals may attempt to conceal their preferences for purposes of taking certain strategic advantages or simply misrepresent their own preferences due to lack of sufficient information [98]. This may lead to preferences that are distinct from the others, resulting in an empty intersection (although it will not affect the union). However, all of the individuals should contribute to the outcome, although one outlier should not affect the overall result. This work therefore extends the original definition of RF sets to a more general version with the use of the OWA operators, which is defined as follows.

Definition 7.1. Let P be an equivalence relation on X and F_l, $l \in \{1, \ldots, J\}$, be FSs in C_o ($C_o \in X$), the LA and UA are a pair of FSs with MFs defined by the following, respectively:

$$\begin{aligned} \mu_{\underline{P}C_o}(x \in [x]_P) &= \text{OWA}\{\mu_{F_l}(x) | x \in [x]_P\} = \sum_{l=1}^{J} w_l \mu_{F_l}(x) \\ \mu_{\overline{P}C_o}(x \in [x]_P) &= \text{OWA}\{\mu_{F_l}(x) | x \in [x]_P\} = \sum_{l=1}^{J} w_l \mu_{F_l}(x) \end{aligned} \tag{7.1}$$

where the weight vector $W = (w_1, w_2, \ldots, w_J)^T$ can be computed using different operators as mentioned before.

Note that when using the Min and Max operators for the calculation of LAs and UAs respectively, the results remain the same as those in Definition 4.2. That is, the original is a specific case of this new definition.

The FSs are aggregated using a partitioning-based method to discretise the input space in this work. The domain of each observed variable x_j, $j = 1, \ldots, M$, is partitioned into a set of discretised values $D_j = \{F_{j1}, \ldots, F_{j|D_j|}\}$, where $|D_j|$ denotes the cardinality of this set. Therefore, given J observations of a variable x_j, the aggregated observation of this variable is calculated using the following OWA operator:

$$F_{\mathrm{OWA}_j} = \sup_{k\in\{0,\ldots,|D_j|\}} \sum_{l=1}^{J} w_{kl}\mu_{F_l}(\min_{x_j} + k * \frac{\max_{x_j} - \min_{x_j}}{|D_j|}) \tag{7.2}$$

where $\max_{x_j}$ and $\min_{x_j}$ are the maximum and minimum values of the jth observed values F_{jl}, $l = 1, \ldots, J$.

7.3 Experiment and Evaluation

A simulated example is used in this section to validate the efficacy of the proposed work. The results obtainable by the proposed AI method are utilised to compare with those by the two IA methods (the proposed and an existing technique).

7.3.1 Experimental Set-up

Individuals may represent their opinions in the form of crisp or fuzzy terms. Occasionally, when only crisp numbers are provided, a fuzzification process is needed. In this simulation-based experimentation, a base function of three crisp input variables, shown in Equation (7.3) is chosen to establish a sparse rule base. A fuzzy rule is generated by fuzzifying the crisp inputs and their associated function output, where a numerical value a is converted to an FS A with a random function f: $A = ((a-f)-f, a-f, a+f, (a+f)+f)$. This provides a simple non-linear (sparse) rule base suitable for the purpose of current investigation.

$$y = 1 + \sqrt{x_1} + \frac{1}{x_2} + \frac{1}{\sqrt{x_3^3}} \tag{7.3}$$

To evaluate the proposed approach, the output y which is computed from the base function, is assumed to be the ground truth for interpolated results. Without losing generality, the arithmetic mean is used for the OWA operator and regarded as the ground truth for the outcome of the aggregation process.

The first comparison is between the proposed AI and IA methods. In this comparison, the extended RF sets are applied to aggregate the derived individuals' solutions in IA and the observed opinions in AI, respectively. The Max operator is selected to calculate LAs, while the DOWA and Clus-DOWA operators are used to compute UAs in order to ensure a purely data-driven implementation. For the sake of reducing computational complexity, the aggregated results are simulated with trapezoidal MFs. The proposed RF interpolation is employed in both IA and AI methods. Thus, two opposite processes are implemented with the proposed approach.

The comparison is also carried out between the proposed AI method and an existing IA method where T-FRI is used for interpolation and the DLSM [182] for aggregation. The weight function used in the existing work of [182] is defined by

$$w_q = \frac{1/\mathrm{Rep}(O_q)}{\sum_{q=1}^{Q} 1/\mathrm{Rep}(O_q)}, \quad q = 1, \ldots, Q \tag{7.4}$$

where $\mathrm{Rep}(O_q)$ is the Rep of the qth computed output value O_q.

In the present simulation-based experimental evaluation, the Reps of the resultant sets of using IA or AI are recorded. They are then compared against their corresponding ground truth calculated using the base function. The range error (RE) and the root-mean-square error (RMSE) are adopted here to analyse the accuracy of the three different approaches:

$$\begin{aligned} \epsilon_{RE} &= \frac{|O_q - G_q|}{\max_y - \min_y}, \quad q = 1, \ldots, Q \\ \epsilon_{RMSE} &= \sqrt{\frac{1}{Q}\sum_{q=1}^{Q}(O_q - G_q)^2} \end{aligned} \tag{7.5}$$

where $\max_y$ and $\min_y$ are the maximum and minimum values of the consequent variable, and O_q and G_q denote the qth computed output value and its corresponding ground truth, respectively.

7.3.2 Results and Discussion

Since stochastic elements are presented in the generation of observations, the evaluation process is repeated 100 times. Tables 7.1, 7.2 and 7.3 list the percentage results of the averaged RE and RMSE, where AN is the number of antecedent variables and O is the number of individual observations. The former two tables show the first comparison with the DOWA operator being used in Table 7.1 and the Clus-DOWA operator in Table 7.2, while the results of using the proposed AI method that obtained in the first comparison are also utilised for the second comparison as listed in Table 7.3.

Table 7.1: Comparison of accuracies - DOWA operator

	Accuracies	AN = 1			AN = 2			AN = 3		
		O = 5	O = 20	O = 50	O = 5	O = 20	O = 50	O = 5	O = 20	O = 50
IA	$\epsilon_{\%RE}$	1.13	0.87	0.82	1.56	1.59	2.06	1.51	1.18	1.19
	$\epsilon_{\%RMSE}$	10.86	8.41	7.58	15.14	15.22	18.59	14.26	11.34	11.02
AI	$\epsilon_{\%RE}$	1.44	0.80	0.59	1.44	0.80	0.57	1.28	0.98	0.62
	$\epsilon_{\%RMSE}$	12.78	7.49	5.37	14.72	7.83	5.29	12.13	10.15	6.01

Table 7.2: Comparison of accuracies - Clus-DOWA operator

		AN = 1			AN = 2			AN = 3		
	Accuracies	O = 5	O = 20	O = 50	O = 5	O = 20	O = 50	O = 5	O = 20	O = 50
IA	$\epsilon_{\%RE}$	0.85	0.78	0.74	1.15	1.28	1.64	1.00	1.06	1.04
	$\epsilon_{\%RMSE}$	7.85	7.28	6.95	10.77	12.33	15.48	10.15	9.97	9.75
AI	$\epsilon_{\%RE}$	1.01	0.79	0.56	1.27	0.72	0.55	1.19	0.95	0.59
	$\epsilon_{\%RMSE}$	9.72	7.27	5.14	12.85	7.04	5.14	11.04	10.20	5.69

Table 7.3: Comparison of accuracies - T-FRI and DLSM

	Accuracies	AN = 1			AN = 2			AN = 3		
		O = 5	O = 20	O = 50	O = 5	O = 20	O = 50	O = 5	O = 20	O = 50
IA	$\epsilon_{\%RE}$	0.71	0.60	0.51	1.10	0.80	0.69	0.79	0.55	0.53
	$\epsilon_{\%RMSE}$	6.45	5.23	4.37	10.32	7.28	5.71	7.44	4.98	4.55
AI (D)	$\epsilon_{\%RE}$	1.44	0.80	0.59	1.44	0.80	0.57	1.28	0.98	0.62
	$\epsilon_{\%RMSE}$	12.78	7.49	5.37	14.72	7.83	5.29	12.13	10.15	6.01
AI (C)	$\epsilon_{\%RE}$	1.01	0.79	0.56	1.27	0.72	0.55	1.19	0.95	0.59
	$\epsilon_{\%RMSE}$	9.72	7.27	5.14	12.85	7.04	5.14	11.04	10.20	5.69

It is obvious that for the first comparison, overall, the AI method outperforms the IA method, especially when the number of observations becomes large. The accuracy of the proposed approach is generally higher than that of its opposite process. This is achieved with less computational complexity (as pointed out previously).

Note that the accuracy attainable by the AI method is not so good as its counterpart in the second comparison when the number of observations is small. However, it is important to point out that the computational overheads of IA is significantly greater than that of AI. Thus, IA may be difficult for particular GDM applications with a larger number of opinions or where a timely generation of solutions is required. This is verified by the result in that the accuracy of AI improves and becomes comparable to that of IA as the number of observations is increased. This implies that the proposed approach is suitable for complex systems in GDM. In addition, the accuracy of using the Clus-DOWA operator is consistently (with just one exception) higher than that of utilising the DOWA operator.

7.4 Summary

This chapter has presented an OWA-based FRI technique for GDM. In order to better represent the underlying uncertainty, the proposed RF set representation has been extended to a more general version with the use of the OWA operator. Also, the extended RF sets are utilised in an GDM problem to evaluate the efficacy of the extended work.

The extended RF set theoretic approach and the proposed T-FRI approach are employed for aggregating individuals' preferences and interpolating the final decision in a purely data-driven manner. According to the simulated experimentation, the proposed technique can reduce the system processing time, while assuring the decision accuracy. This demonstrates that the proposed work is useful for GDM in complex systems.

Chapter 8

Conclusion

THIS chapter concludes the thesis. A summary of the research as detailed in the preceding chapters is presented, with a focus on the main contribution: exploiting the uncertain information in the knowledge for FRI. The thesis has demonstrated that the developed higher order FRI framework has utilised HOFSs effectively for the task of representing and handling uncertainty. The proposed extension to HOFS further enhances the efficacy of the framework. Future developments of the higher order FRI techniques which have been suggested throughout the thesis are enumerated with preliminary suggestions as to how to approach such further work.

8.1 Summary of Thesis

This thesis is concerned primarily with the representation and handling of knowledge with uncertain information in the context of FRI. FRI is a special type of fuzzy inference where the rule bases are sparse. Fuzzy inference was originally proposed in order to handle the inexactness during information processing. Indeed, as a special type of fuzzy inference, FRI not only inherits the properties of fuzzy inference, but also has its own property. That is, FRI is able to deal with inference with an incomplete knowledge base, which is epitomised by the sparse rule bases used in FRI.

However, due to different types of uncertainty involved in FRI, the difficulty in defining the required precise-valued MFs of the FSs significantly restricts the application of conventional FRI techniques. When facing such a higher order uncertainty, a simple approach may be just ignoring this higher-level information. Yet, an obvious drawback of this is that significant information may be lost, resulting in unacceptable inference conclusions. However, the way uncertainty may be represented and processed also depends on the choice of what technique to use. There are different uncertainty representation and handling techniques that may be exploited in devising FRI mechanisms. It is therefore desirable to have a generic framework in which such techniques may be unified and further developed. A higher order framework has been proposed here for both representing the knowledge involving higher order uncertainty and facilitating interpolation and extrapolation with such knowledge.

Before introducing the higher order framework, a thorough review of the existing body of literature on FRI has been given in Chapter 2. In particular, the majority of the existing FRI approaches are categorised into two groups: one-step FRI and two-step FRI. Each group has been examined with a representative approach as well as its extensions and improvements in detail. Besides FRI, basic knowledge representations for characterising different types of uncertainty have also been systematically introduced, including RSs and type-2 FSs. Also, as the basis for the extension of the framework, the OWA and similarity measure operators for information aggregation have been outlined.

The proposed higher order framework is a generalisation of the transformation-based FRI techniques. It aims to offer greater flexibility in handling different types of uncertainty that may be represent in sparse rule bases and observations. Two main components: higher order knowledge representation and higher order rule interpolation have been detailed in Chapter 3. The HOFSs in particular have been designed to capture and represent such uncertainty, in which the lower and upper bounds characterise the range of uncertainty. Then, higher order interpolation has also been designed to perform interpolation and extrapolation in terms of HOFSs.

In order to realise the proposed higher order framework, two implementations have been carried out. First, the implementation based on the use of RF sets has been presented in Chapter 4. Inspired by the concept of RSs, a new definition of RF sets has been proposed in order to establish this implementation, which is characterised by the lower and upper approximation MFs. An algorithm for RF rule interpolation has been subsequently explained assuming that sparse rule bases involving RF variables are available. Then, a proof has been provided in order to verify that the RF approach is indeed compatible with the original T-FRI.

A type-2 FRI technique has been presented as an alternative implementation of the proposed framework in Chapter 5. As an extension to the conventional (type-1) FSs, type-2 FSs are useful for handling uncertainty. The basic concepts involved have been depicted and an approach for type-2 FRI has been presented. Illustrative examples have shown that as with the RF-based implementation this approach is of natural appeal for FRI. A comparison between type-2 FSs and RF sets has been provided. Discussion concerning general type-2 FSs and interval type-2 FSs has also been given, supported with an illustrative example, which has shown different but similar resultant interpolated conclusions between them, both subsuming conventional type-1 FS-based implementation as a specific case.

The effectiveness of the proposed framework has been evaluated in Chapter 6 by employing a practical application of predicting diarrhoeal disease rates in remote villages. Experts have attempted to model how environmental change may influence disease burden so that they can predict the relevant disease rate. However, such modelling can be a great challenge for experts due to the inexactness of the acquired information and the incompleteness of the obtained knowledge. Moreover, different experts may have different kinds of expertise, resulting in similar but different expert rules and observations. The experimental work has shown that such problems can be dealt with the use of the proposed work. Chapter 6 has also included further application of the framework to real datasets concerning decision making problems, supporting the derivation of consensual and consistent decisions.

Finally, a theoretical extension to the proposed RF sets has been described in Chapter 7. The original definition of RF sets has been extended to a more general version with the use of the OWA operator, leading to an OWA-based FRI method that has then been applied to group decision making. This helps ensure that the best decision can be made with respect to the available information and knowledge possessed by all group members. Experimental results have demonstrated the efficacy of the proposed work in producing interpolated results that entail overall decision accuracy.

8.2 Future Work

Although promising, much can be done to further improve the work presented in the thesis. The following subsections address a number of interesting issues whose successful solutions will help to strengthen the current research and approaches.

8.2.1 Short-term Developments

This subsection discusses on extensions and tasks that could be readily implemented if additional time were available.

8.2.1.1 Framework Implementation with Gaussian Membership Functions

One of the significant steps in the higher order framework is defining fuzzy MFs and the corresponding values. There are many choices for the types of MF which may be used, such as triangular, trapezoidal, or Gaussian. The current framework only considers polygonal MFs. In addition, Gaussian MF is another popular method for specifying an FS for two reasons. Firstly, a fuzzy system with Gaussian MF has been shown to be a universal approximator of any non-linear functions on a compact set [177, 178]. Secondly, a multi-dimensional Gaussian MF generated during the learning process can be decomposed into the product of one-dimensional Gaussian MFs [96]. Also, Gaussian MFs are usually preferred for their smooth transition and simple adaptability. A Gaussian MF is entirely specified by the two parameters: the mean and the standard deviation. As such, it would be interesting to implement the framework with the use of Gaussian MFs.

8.2.1.2 Framework Implementation with Weighted Fuzzy Rules

Improving the generalised capability of fuzzy IF-THEN rules extracted from training data is very important for a rule-based fuzzy system [181]. In practice, a priori information may exist about the data pairs [179]. Certain data may be very useful and crucial, but others may be less useful and may even contain misleading information or measurement errors. A degree can therefore be assigned to each data pair that expresses the belief of its usefulness. When taking into account belief degrees in fuzzy rules, the relative weight of each rule among all rules (the rule weight) and the relative weight of each antecedent variable (the antecedent weight), jointly constitute the weight that may be associated with the resulting fuzzy rules. However, the framework presented in this thesis implements fuzzy inference without the use of such rule weight information. It would be helpful to investigate the performance of the framework by learning weights on the underlying constructed rule bases in an effort to maximise the utilisation of uncertain information. Heuristic or evolutionary algorithms [71, 79, 81, 102] could offer a solution for this task.

8.2.1.3 Evaluation on More Realistic Data

The datasets employed in this thesis are public benchmark data, available through the UCI machine learning repository [3]. They have been sourced from real-world problem scenarios. Even though the utilised datasets reflect the high performance of the proposed framework, it would be interesting to evaluate the framework on other real-world problems. Application to such datasets would help to further demonstrate the applicability and versatility of the framework.

8.2.2 Long-term Developments

This subsection proposes two future directions that could each form the basis of a much more significant piece of research.

8.2.2.1 Generalisation of the Framework for Other Interpolation Techniques

The higher order framework proposed in this work presents itself only in terms of the transformation-based interpolation technique, but this is not fundamentally restricted by the underlying framework. Since certain existing higher order FRI methods (e.g., [37, 38, 40, 41]) are based on non-transformation techniques, a more general framework might be useful to incorporate the consideration of different techniques. The generalisation of the development of this framework will substantially improve the applicability of the work. Also, the framework targets uncertainty problems encountered during FRI only. It would be interesting to investigate how this framework can help with conventional fuzzy inference systems given the fact that they also face the issue of dealing with different types of uncertain information. It would be worthwhile developing a unified uncertainty representation and handling platform that implements both conventional fuzzy inference and FRI. This is of great importance since FRI techniques and conventional fuzzy inference may be applied to a single complex problem in order to make inference possible for both dense and sparse rule bases [208].

8.2.2.2 Fuzzy Rule Base Simplification

If the essential information contained in a rule base can be extracted and represented by a subset of the original rules, the new compressed rule base can still be used for calculating approximately the same conclusions [105, 161, 209]. In particular, for fuzzy rule-based models acquired from numerical data, redundancy may be present in the form of similar FSs or rules that represent compatible concepts and their relations. This results in an unnecessarily complex and more opaque linguistic description of the system. Thus, it is potentially very useful to apply the proposed framework to reduce the number of FSs in the rule base through a reverse engineering process. That is, neighbouring rules may be replaced by an interpolate rule [76, 77]. Also, close and similar FSs are merged to create a common HOFS to replace them. If the redundancy in the rule base is high, merging close and similar FSs may result in equal rules that can also be merged, thereby reducing the total number of rules. The reduced rule base may be computationally more efficient and linguistically more interpretable.

Appendix A

Publications Arising from the Thesis

All publications are presented in chronological order.

- C. Chen and Q. Shen. Towards rough-fuzzy rule interpolation. *Proceedings of the 11th UK Workshop on Computational Intelligence*, 2011.
- C. Chen and Q. Shen. A new method for rule interpolation inspired by rough-fuzzy sets. *Proceedings of the 21st International Conference on Fuzzy Systems*. pp. 1098-1105, 2012.
- C. Chen, C. Quek and Q. Shen. Scale and move transformation-based fuzzy rule interpolation with interval type-2 fuzzy sets. *Proceedings of the 22nd International Conference on Fuzzy Systems*. pp. 1-8, 2013.
- C. Chen and Q. Shen. OWA-based fuzzy rule interpolation for group decision making. *Proceedings of the 23rd International Conference on Fuzzy Systems*. pp. 1319-1326, 2014.
- L. Yang, C. Chen, N. Jin, X. Fu and Q. Shen. Closed form fuzzy interpolation with interval type-2 fuzzy sets. *Proceedings of the 23rd International Conference on Fuzzy Systems*. pp. 2184-2191, 2014.
- C. Chen, S. Jin, Y. Li and Q. Shen. Backward rough-fuzzy rule interpolation. To appear in *Proceedings of the 24th International Conference on Fuzzy Systems*.
- C. Chen, N. Mac Parthaláin, C. Quek and Q. Shen. Rough-fuzzy rule interpolation. Under review for journal publication.

Appendix B

List of Acronyms

CCL Chang, Chen, and Liau

CK Chen and Ko

Clus-DOWA Cluster-based DOWA

DLSM Defuzzification-based least squares method

DOWA Dependent OWA

FIS Fuzzy inference system

FLS Fuzzy logic system

FOU Footprint of uncertainty

FRI Fuzzy rule interpolation

FS Fuzzy set

GDM Group decision making

HCL Hsiao, Chen, and Lee

HOFS Higher order fuzzy set

IOWA Induced OWA

KH Kóczy and Hirota

LA Lower approximation

LSDM Least squares distance method

MACI Modified α-cut based interpolation

MF Membership function

OAM Optimal aggregation method

OWA Ordered weighted averaging

QMY Qiao, Mizumoto, and Yan

Rep Representative value

RF Rough-fuzzy

RS Rough set

SAM Similarity aggregation method

SC Soft computing

T-FRI Scale and move transformation-based FRI approach

UA Upper approximation

VKK Vass, Kalmár, and Kóczy

Bibliography

[1] R. Akerkar and P. Sajja, *Knowledge-based systems*. Jones & Bartlett Publishers, 2010.

[2] E. Alpaydin, *Introduction to machine learning*. MIT Press, 2004.

[3] K. Bache and M. Lichman, "UCI machine learning repository," 2013. [Online]. Available: http://archive.ics.uci.edu/ml

[4] T. Back, *Evolutionary algorithms in theory and practice: evolution strategies, evolutionary programming, genetic algorithms*. Oxford University Press, 1996.

[5] M. Banerjee and S. Pal, "Roughness of a fuzzy set," *Information Sciences*, vol. 93, no. 3, pp. 235–246, 1996.

[6] P. Baranyi, L. Kóczy, and T. Gedeon, "A generalized concept for fuzzy rule interpolation," *IEEE Transactions on Fuzzy Systems*, vol. 12, no. 6, pp. 820–837, 2004.

[7] A. Bebbington and T. Perreault, "Social capital, development, and access to resources in highland Ecuador," *Economic Geography*, vol. 75, no. 4, pp. 395–418, 1999.

[8] R. E. Bellman and L. A. Zadeh, "Decision-making in a fuzzy environment," *Management Science*, vol. 17, no. 4, pp. B141–B164, 1970.

[9] J. C. Bezdek, *Pattern recognition with fuzzy objective function algorithms*. Kluwer Academic Publishers, 1981.

[10] C. M. Bishop, *Neural networks for pattern recognition*. Clarendon Press, 1995.

[11] ——, *Pattern recognition and machine learning*. Springer, 2006.

[12] P. P. Bonissone, "Soft computing: the convergence of emerging reasoning technologies," *Soft Computing*, vol. 1, no. 1, pp. 6–18, 1997.

[13] T. Boongoen and Q. Shen, "Clus-DOWA: a new dependent OWA operator," in *Proceedings of the IEEE International Conference on Fuzzy Systems*, 2008, pp. 1057–1063.

[14] ——, "Nearest-neighbor guided evaluation of data reliability and its applications," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 40, no. 6, pp. 1622–1633, 2010.

[15] B. Bouchon-Meunier and L. Valverde, "A fuzzy approach to analogical reasoning," *Soft Computing*, vol. 3, no. 3, pp. 141–147, 1999.

[16] R. J. Brachman and H. J. Levesque, *Readings in knowledge representation*. Morgan Kaufmann Publishers Inc., 1985.

[17] J. J. Buckley, “Sugeno type controllers are universal controllers,” *Fuzzy Sets and Systems*, vol. 53, no. 3, pp. 299–303, 1993.

[18] J. Buckley, W. Siler, and D. Tucker, “A fuzzy expert system,” *Fuzzy Sets and Systems*, vol. 20, no. 1, pp. 1–16, 1986.

[19] H. Bustince, “Indicator of inclusion grade for interval-valued fuzzy sets. Application to approximate reasoning based on interval-valued fuzzy sets,” *International Journal of Approximate Reasoning*, vol. 23, no. 3, pp. 137–209, 2000.

[20] C. Carlsson and R. Fullér, “Fuzzy multiple criteria decision making: recent developments,” *Fuzzy Sets and Systems*, vol. 78, no. 2, pp. 139–153, 1996.

[21] J. L. Castro, “Fuzzy logic controllers are universal approximators,” *IEEE Transactions on Systems, Man and Cybernetics*, vol. 25, no. 4, pp. 629–635, 1995.

[22] S. Chanas and D. Kuchta, “Further remarks on the relation between rough and fuzzy sets,” *Fuzzy Sets and Systems*, vol. 47, no. 3, pp. 391–394, 1992.

[23] Y. Chang, S. Chen, and C. Liau, “Fuzzy interpolative reasoning for sparse fuzzy-rule-based systems based on the areas of fuzzy sets,” *IEEE Transactions on Fuzzy Systems*, vol. 16, no. 5, pp. 1285–1301, 2008.

[24] A. Charnes, W. W. Cooper, and E. Rhodes, “Measuring the efficiency of decision making units,” *European Journal of Operational Research*, vol. 2, no. 6, pp. 429–444, 1978.

[25] C. Chen and Q. Shen, “Towards rough-fuzzy rule interpolation,” in *Proceedings of the UK Workshop on Computational Intelligence*, 2011.

[26] ——, “A new method for rule interpolation inspired by rough-fuzzy sets,” in *Proceedings of the IEEE International Conference on Fuzzy Systems*, 2012, pp. 1098–1105.

[27] C. Chen, C. Quek, and Q. Shen, “Scale and move transformation-based fuzzy rule interpolation with interval type-2 fuzzy sets,” in *Proceedings of the IEEE International Conference on Fuzzy Systems*, 2013, pp. 1–8.

[28] C. Chen and Q. Shen, “Owa-based fuzzy rule interpolation for group decision making,” in *Proceedings of the IEEE International Conference on Fuzzy Systems*, 2014, pp. 1319–1326.

[29] S.-M. Chen, Y.-C. Chang, and J.-S. Pan, “Fuzzy rules interpolation for sparse fuzzy rule-based systems based on interval type-2 Gaussian fuzzy sets and genetic algorithms,” *IEEE Transactions on Fuzzy Systems*, vol. 21, no. 3, pp. 412–425, 2013.

[30] S.-J. Chen and C.-L. Hwang, *Fuzzy multiple attribute decision making: methods and applications*. Springer, 1992.

[31] S.-M. Chen, “A new approach to handling fuzzy decisionmaking problems,” *IEEE Transactions on Systems, Man and Cybernetics*, vol. 18, no. 6, pp. 1012–1016, 1988.

[32] ——, "A weighted fuzzy reasoning algorithm for medical diagnosis," *Decision Support Systems*, vol. 11, no. 1, pp. 37–43, 1994.

[33] ——, "Aggregating fuzzy opinions in the group decision-making environment," *Cybernetics & Systems*, vol. 29, no. 4, pp. 363–376, 1998.

[34] S.-M. Chen and Y.-C. Chang, "Weighted fuzzy rule interpolation based on GA-based weight-learning techniques," *IEEE Transactions on Fuzzy Systems*, vol. 19, no. 4, pp. 729–744, 2011.

[35] S.-M. Chen, Y.-K. Ko, Y.-C. Chang, and J.-S. Pan, "Weighted fuzzy interpolative reasoning based on weighted increment transformation and weighted ratio transformation techniques," *IEEE Transactions on Fuzzy Systems*, vol. 17, no. 6, pp. 1412–1427, 2009.

[36] S. Chen and Y. Chang, "Fuzzy rule interpolation based on principle membership functions and uncertainty grade functions of interval type-2 fuzzy sets," *Expert Systems with Applications*, vol. 38, no. 9, pp. 11 573–11 580, 2011.

[37] ——, "Fuzzy rule interpolation based on the ratio of fuzziness of interval type-2 fuzzy sets," *Expert Systems with Applications*, vol. 38, no. 10, pp. 12 202–12 213, 2011.

[38] S. Chen, Y. Chang, and J. Pan, "Fuzzy rules interpolation for sparse fuzzy rule-based systems based on interval type-2 Gaussian fuzzy sets and genetic algorithms," *IEEE Transactions on Fuzzy Systems*, vol. 21, no. 3, pp. 412–425, 2013.

[39] S. Chen and Y. Ko, "Fuzzy interpolative reasoning for sparse fuzzy rule-based systems based on α-cuts and transformations techniques," *IEEE Transactions on Fuzzy Systems*, vol. 16, no. 6, pp. 1626–1648, 2008.

[40] S. Chen and L. Lee, "Fuzzy interpolative reasoning for sparse fuzzy rule-based systems based on interval type-2 fuzzy sets," *Expert Systems with Applications*, vol. 38, no. 8, pp. 9947–9957, 2011.

[41] S. Chen, L. Lee, and V. R. Shen, "Weighted fuzzy interpolative reasoning systems based on interval type-2 fuzzy sets," *Information Sciences*, vol. 248, pp. 15–30, 2013.

[42] R. R. Colwell, P. R. Epstein, D. Gubler, N. Maynard, A. J. McMichael, J. A. Patz, R. B. Sack, and R. Shope, "Climate change and human health," *Science*, vol. 279, no. 5353, pp. 968–969, 1998.

[43] O. Cordón, F. Herrera, and P. Villar, "Generating the knowledge base of a fuzzy rule-based system by the genetic learning of the data base," *IEEE Transactions on Fuzzy Systems*, vol. 9, no. 4, pp. 667–674, 2001.

[44] S. Coupland and R. John, "Geometric type-1 and type-2 fuzzy logic systems," *IEEE Transactions on Fuzzy Systems*, vol. 15, no. 1, pp. 3–15, 2007.

[45] R. Davis and D. B. Lenat, *Knowledge-based systems in artificial intelligence*. McGraw-Hill, 1982.

[46] K. Deb, *Multi-objective optimization using evolutionary algorithms*. John Wiley & Sons, 2001.

[47] M. Detyniecki, "Fundamentals on aggregation operators," *Berkeley Initiative in Soft Computing, Computer Science Division*, 2001.

[48] R. Diao, F. Chao, T. Peng, N. Snooke, and Q. Shen, "Feature selection inspired classifier ensemble reduction," *IEEE Transactions on Cybernetics*, vol. 44, no. 8, pp. 1259–1268, 2014.

[49] A. Driankov and R. Unberhauen, *An introduction to fuzzy control*. Springer-Verlag, 1993.

[50] D. Dubois and H. Prade, "Rough fuzzy sets and fuzzy rough sets," *International Journal of General Systems*, vol. 17, no. 2, pp. 191–209, 1990.

[51] ——, "Putting rough sets and fuzzy sets together," *Intelligent Decision Support*, pp. 203–232, 1992.

[52] ——, "What are fuzzy rules and how to use them," *Fuzzy Sets and Systems*, vol. 84, no. 2, pp. 169–185, 1996.

[53] ——, "Interval-valued fuzzy sets, possibility theory and imprecise probability," in *Proceedings of the International Conference in Fuzzy Logic and Technology*, 2005.

[54] D. DuBois and H. Prade, *Fuzzy sets and systems: theory and applications*. Academic Press, 1980.

[55] D. Dubois and H. Prade, "A review of fuzzy set aggregation connectives," *Information Sciences*, vol. 36, no. 1, pp. 85–121, 1985.

[56] ——, "Gradual inference rules in approximate reasoning," *Information Sciences*, vol. 61, no. 1, pp. 103–122, 1992.

[57] M. B. Eisen, P. T. Spellman, P. O. Brown, and D. Botstein, "Cluster analysis and display of genome-wide expression patterns," *Proceedings of the National Academy of Sciences*, vol. 95, no. 25, pp. 14 863–14 868, 1998.

[58] J. N. Eisenberg, W. Cevallos, K. Ponce, K. Levy, S. J. Bates, J. C. Scott, A. Hubbard, N. Vieira, P. Endara, M. Espinel *et al.*, "Environmental change and infectious disease: how new roads affect the transmission of diarrheal pathogens in rural Ecuador," *Proceedings of the National Academy of Sciences*, vol. 103, no. 51, pp. 19 460–19 465, 2006.

[59] D. Filev and R. R. Yager, "On the issue of obtaining OWA operator weights," *Fuzzy Sets and Systems*, vol. 94, no. 2, pp. 157–169, 1998.

[60] X. Fu and Q. Shen, "Fuzzy compositional modeling," *IEEE Transactions on Fuzzy Systems*, vol. 18, no. 4, pp. 823–840, 2010.

[61] J. C. Giarratano and G. Riley, *Expert systems: principles and programming*. Brooks/Cole Publishing Co., 1989.

[62] A. Gordon and M. Vichi, "Fuzzy partition models for fitting a set of partitions," *Psychometrika*, vol. 66, no. 2, pp. 229–247, 2001.

[63] M. B. Gorzałczany, "A method of inference in approximate reasoning based on interval-valued fuzzy sets," *Fuzzy Sets and Systems*, vol. 21, no. 1, pp. 1–17, 1987.

[64] I. Graham and P. L. Jones, *Expert systems: knowledge, uncertainty, and decision*. Chapman & Hall, 1988.

[65] C. Grootaert and T. Van Bastelaer, *Understanding and measuring social capital: a multidisciplinary tool for practitioners*. World Bank, 2002.

[66] J. Grzymala-Busse, "Knowledge acquisition under uncertainty–a rough set approach," *Journal of Intelligent and Robotic Systems*, vol. 1, no. 1, pp. 3–16, 1988.

[67] S. Guillaume, "Designing fuzzy inference systems from data: an interpretability-oriented review," *IEEE Transactions on Fuzzy Systems*, vol. 9, no. 3, pp. 426–443, 2001.

[68] S. Haykin, *Neural networks: a comprehensive foundation*. Prentice Hall, 1994.

[69] F. Herrera and E. Herrera-Viedma, "A model of consensus in group decision making under linguistic assessments," *Fuzzy Sets and Systems*, vol. 78, no. 1, pp. 73–87, 1996.

[70] ——, "Linguistic decision analysis: steps for solving decision problems under linguistic information," *Fuzzy Sets and Systems*, vol. 115, no. 1, pp. 67–82, 2000.

[71] J. H. Holland, *Adaptation in natural and artificial systems*. University of Michigan Press, 1975.

[72] T.-P. Hong and C.-Y. Lee, "Induction of fuzzy rules and membership functions from training examples," *Fuzzy Sets and Systems*, vol. 84, no. 1, pp. 33–47, 1996.

[73] W. Hsiao, S. Chen, and C. Lee, "A new interpolative reasoning method in sparse rule-based systems," *Fuzzy Sets and Systems*, vol. 93, no. 1, pp. 17–22, 1998.

[74] H.-M. Hsu and C.-T. Chen, "Aggregation of fuzzy opinions under group decision making," *Fuzzy Sets and Systems*, vol. 79, no. 3, pp. 279–285, 1996.

[75] Z. Huang and Q. Shen, "A new fuzzy interpolative reasoning method based on center of gravity," in *Proccedings of the IEEE International Conference on Fuzzy Systems*, vol. 1, 2003, pp. 25–30.

[76] ——, "Fuzzy interpolative reasoning via scale and move transformations," *IEEE Transactions on Fuzzy Systems*, vol. 14, no. 2, pp. 340–359, 2006.

[77] ——, "Fuzzy interpolation and extrapolation: a practical approach," *IEEE Transactions on Fuzzy Systems*, vol. 16, no. 1, pp. 13–28, 2008.

[78] V.-N. Huynh, Y. Nakamori, and J. Lawry, "A probability-based approach to comparison of fuzzy numbers and applications to target-oriented decision making," *IEEE Transactions on Fuzzy Systems*, vol. 16, no. 2, pp. 371–387, 2008.

[79] H. Ishibuchi and T. Nakashima, "Effect of rule weights in fuzzy rule-based classification systems," *IEEE Transactions on Fuzzy Systems*, vol. 9, no. 4, pp. 506–515, 2001.

[80] H. Ishibuchi, K. Nozaki, N. Yamamoto, and H. Tanaka, “Selecting fuzzy if-then rules for classification problems using genetic algorithms,” *IEEE Transactions on Fuzzy Systems*, vol. 3, no. 3, pp. 260–270, 1995.

[81] H. Ishibuchi and T. Yamamoto, “Rule weight specification in fuzzy rule-based classification systems,” *IEEE Transactions on Fuzzy Systems*, vol. 13, no. 4, pp. 428–435, 2005.

[82] A. K. Jain, *Fundamentals of digital image processing*. Prentice Hall, 1989.

[83] I. L. Janis and L. Mann, *Decision making: a psychological analysis of conflict, choice, and commitment*. New York: Free Press, 1977.

[84] S. Jenei, “Interpolation and extrapolation of fuzzy quantities revisited–an axiomatic approach,” *Soft Computing*, vol. 5, no. 3, pp. 179–193, 2001.

[85] S. Jenei, E.-P. Klement, and R. Konzel, “Interpolation and extrapolation of fuzzy quantities–the multiple-dimensional case,” *Soft Computing*, vol. 6, no. 3, pp. 258–270, 2002.

[86] R. Jensen and Q. Shen, “Fuzzy-rough attribute reduction with application to web categorization,” *Fuzzy Sets and Systems*, vol. 141, no. 3, pp. 469–485, 2004.

[87] ——, “Semantics-preserving dimensionality reduction: rough and fuzzy-rough-based approaches,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, pp. 1457–1471, 2004.

[88] ——, *Computational intelligence and feature selection: rough and fuzzy approaches*. Wiley-IEEE Press, 2008.

[89] ——, “New approaches to fuzzy-rough feature selection,” *IEEE Transactions on Fuzzy Systems*, vol. 17, no. 4, pp. 824–838, 2009.

[90] Z. C. Johanyák and S. Kovács, “A brief survey and comparison on various interpolation based fuzzy reasoning methods,” *Acta Polytechnica Hungarica*, vol. 3, no. 1, pp. 91–105, 2006.

[91] ——, “Fuzzy rule interpolation based on polar cuts,” *Computational Intelligence, Theory and Applications*, pp. 499–511, 2006.

[92] ——, “Survey on various interpolation based fuzzy reasoning methods,” *Production Systems and Information Engineering*, vol. 3, pp. 39–56, 2006.

[93] Z. C. Johanyák, D. Tikk, S. Kovács, and K. W. Wong, “Fuzzy rule interpolation Matlab toolbox–FRI toolbox,” in *Proceedings of the IEEE International Conference on Fuzzy Systems*, 2006, pp. 351–357.

[94] R. John and S. Coupland, “Type-2 fuzzy logic: a historical view,” *Computational Intelligence Magazine*, vol. 2, no. 1, pp. 57–62, 2007.

[95] Y. Joo, H. Hwang, K. Kim, and K. Woo, “Fuzzy system modeling by fuzzy partition and GA hybrid schemes,” *Fuzzy Sets and Systems*, vol. 86, no. 3, pp. 279–288, 1997.

[96] C.-F. Juang and C.-T. Lin, "An online self-constructing neural fuzzy inference network and its applications," *IEEE Transactions on Fuzzy Systems*, vol. 6, no. 1, pp. 12–32, 1998.

[97] J. Kacprzyk, "Group decision making with a fuzzy linguistic majority," *Fuzzy Sets and Systems*, vol. 18, no. 2, pp. 105–118, 1986.

[98] M. F. Kaplan and C. E. Miller, "Group decision making and normative versus informational influence: effects of type of issue and assigned decision rule," *Journal of Personality and Social Psychology*, vol. 53, no. 2, p. 306, 1987.

[99] N. N. Karnik and J. M. Mendel, "Centroid of a type-2 fuzzy set," *Information Sciences*, vol. 132, no. 1, pp. 195–220, 2001.

[100] N. Karnik and J. Mendel, "Introduction to type-2 fuzzy logic systems," in *Proceedings of the IEEE International Conference on Fuzzy Systems*, vol. 2, 1998, pp. 915–920.

[101] N. Karnik, J. Mendel, and Q. Liang, "Type-2 fuzzy logic systems," *IEEE Transactions on Fuzzy Systems*, vol. 7, no. 6, pp. 643–658, 1999.

[102] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of the IEEE International Conference on Neural Networks*, vol. 4, 1995, pp. 1942–1948.

[103] F. Klawonn and J. L. Castro, "Similarity in fuzzy reasoning," *Mathware & Soft Computing*, vol. 2, no. 3, pp. 197–228, 2008.

[104] G. Klir and B. Yuan, *Fuzzy sets and fuzzy logic: theory and applications*. Prentice Hall, 1995.

[105] L. T. Koczy, "Fuzzy if... then rule models and their transformation into one another," *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, vol. 26, no. 5, pp. 621–637, 1996.

[106] L. Kóczy and K. Hirota, "Approximate reasoning by linear rule interpolation and general approximation," *International Journal of Approximate Reasoning*, vol. 9, no. 3, pp. 197–225, 1993.

[107] ——, "Interpolative reasoning with insufficient evidence in sparse fuzzy rule bases," *Information Sciences*, vol. 71, no. 1, pp. 169–201, 1993.

[108] ——, "Size reduction by interpolation in fuzzy rule bases," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 27, no. 1, pp. 14–25, 1997.

[109] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proceedings of the International Joint Conference on Artficial Intelligence*, 1995, pp. 1137–1143.

[110] J. Komorowski, Z. Pawlak, L. Polkowski, and A. Skowron, "Rough sets: a tutorial," *Rough Fuzzy Hybridization: A New Trend in Decision-Making*, pp. 3–98, 1999.

[111] S. Kovács, "Extending the fuzzy rule interpolation "FIVE" by fuzzy observation," *Computational Intelligence, Theory and Applications*, pp. 485–497, 2006.

[112] M. Larsen *et al.*, "Industrial applications of fuzzy logic control," *International Journal of Man-Machine Studies*, vol. 12, no. 1, pp. 3–10, 1980.

[113] J. Lawry, "A methodology for computing with words," *International Journal of Approximate Reasoning*, vol. 28, no. 2, pp. 51–89, 2001.

[114] ——, "A framework for linguistic modelling," *Artificial Intelligence*, vol. 155, no. 1, pp. 1–39, 2004.

[115] ——, *Modelling and reasoning with vague concepts*. Springer, 2006.

[116] ——, "Appropriateness measures: an uncertainty model for vague concepts," *Synthese*, vol. 161, no. 2, pp. 255–269, 2008.

[117] J. Lawry and Y. Tang, "Uncertainty modelling for vague concepts: a prototype theory approach," *Artificial Intelligence*, vol. 173, no. 18, pp. 1539–1558, 2009.

[118] C.-C. Lee, "Fuzzy logic in control systems: fuzzy logic controller–I," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 20, no. 2, pp. 419–435, 1990.

[119] H.-S. Lee, "An optimal aggregation method for fuzzy opinions of group decision," in *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, vol. 3, 1999, pp. 314–319.

[120] ——, "Optimal consensus of fuzzy opinions under group decision making environment," *Fuzzy Sets and Systems*, vol. 132, no. 3, pp. 303–315, 2002.

[121] L. Lee and S. Chen, "Fuzzy interpolative reasoning using interval type-2 fuzzy sets," *New Frontiers in Applied Artificial Intelligence*, pp. 92–101, 2008.

[122] Q. Liang and J. Mendel, "Interval type-2 fuzzy logic systems: theory and design," *IEEE Transactions on Fuzzy Systems*, vol. 8, no. 5, pp. 535–550, 2000.

[123] Y. Lin, G. A. Cunningham III, and S. V. Coggeshall, "Using fuzzy partitions to create fuzzy systems from input-output data and set the initial weights in a fuzzy neural network," *IEEE Transactions on Fuzzy Systems*, vol. 5, no. 4, pp. 614–621, 1997.

[124] D. V. Lindley and D. V. Lindley, *Making decisions*. Wiley New York, 1985.

[125] F. Liu, "An efficient centroid type-reduction strategy for general type-2 fuzzy logic system," *Information Sciences*, vol. 178, no. 9, pp. 2224–2236, 2008.

[126] J. Liu, F. T. Chan, Y. Li, Y. Zhang, and Y. Deng, "A new optimal consensus method with minimum cost in fuzzy group decision," *Knowledge-Based Systems*, vol. 35, pp. 357–360, 2012.

[127] J. Lu, G. Zhang, and D. Ruan, "Intelligent multi-criteria fuzzy group decision-making for situation assessments," *Soft Computing*, vol. 12, no. 3, pp. 289–299, 2008.

[128] X.-J. Ma, Z.-Q. Sun, and Y.-Y. He, "Analysis and design of fuzzy controller and fuzzy observer," *IEEE Transactions on Fuzzy Systems*, vol. 6, no. 1, pp. 41–51, 1998.

[129] E. H. Mamdani, "Application of fuzzy logic to approximate reasoning using linguistic synthesis," *IEEE Transactions on Computers*, vol. 26, pp. 1182–1191, 1977.

[130] E. H. Mamdani and S. Assilian, "An experiment in linguistic synthesis with a fuzzy logic controller," *International Journal of Man-Machine Studies*, vol. 7, no. 1, pp. 1–13, 1975.

[131] J. M. Mendel, L. A. Zadeh, E. Trillas, R. Yager, J. Lawry, H. Hagras, and S. Guadarrama, "What computing with words means to me," *IEEE Computational Intelligence Magazine*, vol. 5, no. 1, pp. 20–26, 2010.

[132] J. Mendel, "Fuzzy logic systems for engineering: a tutorial," *Proceedings of the IEEE*, vol. 83, no. 3, pp. 345–377, 1995.

[133] ——, "Type-2 fuzzy sets: some questions and answers," *IEEE Connections, Newsletter of the IEEE Neural Networks Society*, vol. 1, pp. 10–13, 2003.

[134] ——, "Computing with words and its relationships with fuzzistics," *Information Sciences*, vol. 177, no. 4, pp. 988–1006, 2007.

[135] ——, "Type-2 fuzzy sets and systems: an overview," *Computational Intelligence Magazine*, vol. 2, no. 1, pp. 20–29, 2007.

[136] J. Mendel, R. John, and F. Liu, "Interval type-2 fuzzy logic systems made simple," *IEEE Transactions on Fuzzy Systems*, vol. 14, no. 6, pp. 808–821, 2006.

[137] J. Mendel and R. John, "Type-2 fuzzy sets made simple," *IEEE Transactions on Fuzzy Systems*, vol. 10, no. 2, pp. 117–127, 2002.

[138] J. M. Merigó and A. M. Gil-Lafuente, "The induced generalized OWA operator," *Information Sciences*, vol. 179, no. 6, pp. 729–741, 2009.

[139] R. S. Michalski, J. G. Carbonell, and T. M. Mitchell, *Machine learning: an artificial intelligence approach*. Morgan Kaufmann, 1986.

[140] H. Mitchell and D. Estrakh, "A modified OWA operator and its use in lossless DPCM image compression," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 5, no. 04, pp. 429–436, 1997.

[141] T. M. Mitchell, *Machine learning*. McGraw-Hill, 1997.

[142] S. Mitra, S. K. Pal, and P. Mitra, "Data mining in soft computing framework: a survey," *IEEE Transactions on Neural Networks*, vol. 13, no. 1, pp. 3–14, 2002.

[143] M. Mizumoto and H. Zimmermann, "Comparison of fuzzy reasoning methods," *Fuzzy Sets and Systems*, vol. 8, no. 3, pp. 253–283, 1982.

[144] S. S. Morse, "Factors in the emergence of infectious diseases," *Emerging Infectious Diseases*, vol. 1, no. 1, pp. 7–15, 1995.

[145] R. E. Neapolitan, *Probabilistic reasoning in expert systems: theory and algorithms*. John Wiley & Sons, 1990.

[146] K. Nozaki, H. Ishibuchi, and H. Tanaka, "A simple but powerful heuristic method for generating fuzzy rules from numerical data," *Fuzzy Sets and Systems*, vol. 86, no. 3, pp. 251–270, 1997.

[147] C. P. Pappis and N. I. Karacapilidis, "A comparative assessment of measures of similarity of fuzzy values," *Fuzzy Sets and Systems*, vol. 56, no. 2, pp. 171–174, 1993.

[148] N. Parthaláin, Q. Shen, and R. Jensen, "A distance measure approach to exploring the rough set boundary region for attribute reduction," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 3, pp. 305–317, 2010.

[149] Z. Pawlak, "Rough set theory and its applications," *Journal of Telecommunications and Information Technology*, vol. 3, pp. 7–10, 2002.

[150] ——, "Rough sets and intelligent data analysis," *Information Sciences*, vol. 147, no. 1, pp. 1–12, 2002.

[151] ——, "Rough sets," *International Journal of Computer & Information Sciences*, vol. 11, no. 5, pp. 341–356, 1982.

[152] ——, *Rough sets: theoretical aspects of reasoning about data*. Kluwer Academic Publishers, 1991.

[153] Z. Pawlak, J. Grzymala-Busse, R. Slowinski, and W. Ziarko, "Rough sets," *Communications of the ACM*, vol. 38, no. 11, pp. 88–95, 1995.

[154] Z. Pawlak and A. Skowron, "Rough membership functions," *Advances in the Dempster-Shafer Theory of Evidence*, pp. 251–271, 1994.

[155] J. Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, 1988.

[156] I. Perfilieva, D. Dubois, H. Prade, F. Esteva, L. Godo, and P. Hoďáková, "Interpolation of fuzzy data: analytical approach and overview," *Fuzzy Sets and Systems*, vol. 192, pp. 134–158, 2012.

[157] W. Qiao, M. Mizumoto, and S. Yan, "An improvement to Kóczy and Hirota's interpolative reasoning in sparse fuzzy rule bases," *International Journal of Approximate Reasoning*, vol. 15, no. 3, pp. 185–201, 1996.

[158] J. A. Robinson, "A machine-oriented logic based on the resolution principle," *Journal of the ACM*, vol. 12, no. 1, pp. 23–41, 1965.

[159] M. Roubens, "Fuzzy sets and decision analysis," *Fuzzy Sets and Systems*, vol. 90, no. 2, pp. 199–206, 1997.

[160] J. C. Russ and R. P. Woods, "The image processing handbook," *Journal of Computer Assisted Tomography*, vol. 19, no. 6, pp. 979–981, 1995.

[161] M. Setnes, R. Babuska, U. Kaymak, and H. R. van Nauta Lemke, "Similarity measures in fuzzy rule base simplification," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 28, no. 3, pp. 376–386, 1998.

[162] J. Shao, “Linear model selection by cross-validation,” *Journal of the American statistical Association*, vol. 88, no. 422, pp. 486–494, 1993.

[163] Q. Shen and L. Yang, “Generalisation of scale and move transformation-based fuzzy interpolation,” *Advanced Computational Intelligence and Intelligent Informatics*, vol. 15, no. 3, pp. 288–298, 2011.

[164] Y. Shi and M. Mizumoto, “Some considerations on Kóczy’s fuzzy interpolative reasoning method,” in *Proceedings of the IEEE International Conference on Fuzzy Systems*, vol. 4, 1995, pp. 2117–2122.

[165] ——, “Reasoning conditions on Kóczy’s interpolative reasoning method in sparse fuzzy rule bases. Part II,” *Fuzzy Sets and Systems*, vol. 87, no. 1, pp. 47–56, 1997.

[166] J. F. Sowa, *Knowledge representation: logical, philosophical, and computational foundations*. Course Technology, 1999.

[167] M. Sugeno and G. Kang, “Structure identification of fuzzy model,” *Fuzzy Sets and Systems*, vol. 28, no. 1, pp. 15–33, 1988.

[168] T. Takagi and M. Sugeno, “Fuzzy identification of systems and its applications to modeling and control,” *IEEE Transactions on Systems, Man and Cybernetics*, no. 1, pp. 116–132, 1985.

[169] D. Tikk and P. Baranyi, “Comprehensive analysis of a new fuzzy rule interpolation method,” *IEEE Transactions on Fuzzy Systems*, vol. 8, no. 3, pp. 281–296, 2000.

[170] V. Torra, “The weighted OWA operator,” *International Journal of Intelligent Systems*, vol. 12, no. 2, pp. 153–166, 1997.

[171] T. Tsabadze, “A method for fuzzy aggregation based on group expert evaluations,” *Fuzzy Sets and Systems*, vol. 157, no. 10, pp. 1346–1361, 2006.

[172] I. Turksen, “Interval valued fuzzy sets based on normal forms,” *Fuzzy Sets and Systems*, vol. 20, no. 2, pp. 191–210, 1986.

[173] I. Turksen and Z. Zhong, “An approximate analogical reasoning approach based on similarity measures,” *IEEE Transactions on Systems, Man and Cybernetics*, vol. 18, no. 6, pp. 1049–1056, 1988.

[174] ——, “An approximate analogical reasoning schema based on similarity measures and interval-valued fuzzy sets,” *Fuzzy Sets and Systems*, vol. 34, no. 3, pp. 323–346, 1990.

[175] J. Vaníček, I. Vrana, and S. Aly, “Fuzzy aggregation and averaging for group decision making: a generalization and survey,” *Knowledge-Based Systems*, vol. 22, no. 1, pp. 79–84, 2009.

[176] G. Vass, L. Kalmár, and L. Kóczy, “Extension of the fuzzy rule interpolation method,” in *Proccedings of the International Conference on Fuzzy Stes Theory and Applications*, 1992, pp. 1–6.

[177] L.-X. Wang, “Fuzzy systems are universal approximators,” in *Proceedings of the IEEE International Conference on Fuzzy Systems*, 1992, pp. 1163–1170.

[178] L.-X. Wang and J. M. Mendel, “Fuzzy basis functions, universal approximation, and orthogonal least-squares learning,” *IEEE Transactions on Neural Networks*, vol. 3, no. 5, pp. 807–814, 1992.

[179] ——, “Generating fuzzy rules by learning from examples,” *IEEE Transactions on Systems, Man and Cybernetics*, vol. 22, no. 6, pp. 1414–1427, 1992.

[180] L.-X. Wang, *A course in fuzzy systems*. Prentice Hall, 1999.

[181] X.-Z. Wang and C.-R. Dong, “Improving generalization of fuzzy if-then rules by maximizing fuzzy entropy,” *IEEE Transactions on Fuzzy Systems*, vol. 17, no. 3, pp. 556–567, 2009.

[182] Y.-M. Wang and C. Parkan, “Two new approaches for assessing the weights of fuzzy opinions in group decision analysis,” *Information Sciences*, vol. 176, no. 23, pp. 3538–3555, 2006.

[183] K. Wong, D. Tikk, T. Gedeon, and L. Kóczy, “Fuzzy rule interpolation for multidimensional input spaces with applications: a case study,” *IEEE Transactions on Fuzzy Systems*, vol. 13, no. 6, pp. 809–819, 2005.

[184] D. Wu and J. Mendel, “Uncertainty measures for interval type-2 fuzzy sets,” *Information Sciences*, vol. 177, no. 23, pp. 5378–5393, 2007.

[185] ——, “A comparative study of ranking methods, similarity measures and uncertainty measures for interval type-2 fuzzy sets,” *Information Sciences*, vol. 179, no. 8, pp. 1169–1192, 2009.

[186] Z. Xu, “An overview of methods for determining OWA weights,” *International Journal of Intelligent Systems*, vol. 20, no. 8, pp. 843–865, 2005.

[187] ——, “Dependent OWA operators,” *Modeling Decisions for Artificial Intelligence*, pp. 172–178, 2006.

[188] Z. Xu and Q. Da, “The ordered weighted geometric averaging operators,” *International Journal of Intelligent Systems*, vol. 17, no. 7, pp. 709–716, 2002.

[189] ——, “The uncertain OWA operator,” *International Journal of Intelligent Systems*, vol. 17, no. 6, pp. 569–575, 2002.

[190] ——, “An overview of operators for aggregating information,” *International Journal of Intelligent Systems*, vol. 18, no. 9, pp. 953–969, 2003.

[191] R. R. Yager, “On ordered weighted averaging aggregation operators in multicriteria decisionmaking,” *IEEE Transactions on Systems, Man and Cybernetics*, vol. 18, no. 1, pp. 183–190, 1988.

[192] ——, “Families of OWA operators,” *Fuzzy Sets and Systems*, vol. 59, no. 2, pp. 125–148, 1993.

[193] ——, "A general approach to criteria aggregation using fuzzy measures," *International Journal of Man-Machine Studies*, vol. 39, no. 2, pp. 187–213, 1993.

[194] ——, "Aggregation operators and fuzzy systems modeling," *Fuzzy Sets and Systems*, vol. 67, no. 2, pp. 129–145, 1994.

[195] ——, "Quantifier guided aggregation using OWA operators," *International Journal of Intelligent Systems*, vol. 11, no. 1, pp. 49–73, 1996.

[196] ——, "Generalized OWA aggregation operators," *Fuzzy Optimization and Decision Making*, vol. 3, no. 1, pp. 93–107, 2004.

[197] ——, "Centered OWA operators," *Soft Computing*, vol. 11, no. 7, pp. 631–639, 2007.

[198] R. R. Yager and D. P. Filev, "Generalizing the modeling of fuzzy logic controllers by parameterized aggregation operators," *Fuzzy Sets and Systems*, vol. 70, no. 2, pp. 303–313, 1995.

[199] ——, "Induced ordered weighted averaging operators," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 29, no. 2, pp. 141–150, 1999.

[200] R. R. Yager, D. P. Filev, and T. Sadeghi, "Analysis of flexible structured fuzzy logic controllers," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 24, no. 7, pp. 1035–1043, 1994.

[201] R. R. Yager, L. S. Goldstein, and E. Mendels, "FUZMAR: an approach to aggregating market research data based on fuzzy reasoning," *Fuzzy Sets and Systems*, vol. 68, no. 1, pp. 1–11, 1994.

[202] R. R. Yager and L. A. Zadeh, *Fuzzy sets, neural networks and soft computing*. John Wiley & Sons, 1994.

[203] Y. Yam and L. T. Kóczy, "Representing membership functions as points in high-dimensional spaces for fuzzy interpolation and extrapolation," *IEEE Transactions on Fuzzy Systems*, vol. 8, no. 6, pp. 761–772, 2000.

[204] Y. Yam, M. L. Wong, and P. Baranyi, "Interpolation with function space representation of membership functions," *IEEE Transactions on Fuzzy Systems*, vol. 14, no. 3, pp. 398–411, 2006.

[205] S. Yan, M. Mizumoto, and W. Qiao, "Reasoning conditions on Kóczy's interpolative reasoning method in sparse fuzzy rule bases," *Fuzzy Sets and Systems*, vol. 75, no. 1, pp. 63–71, 1995.

[206] L. Yang and Q. Shen, "Adaptive fuzzy interpolation," *IEEE Transactions on Fuzzy Systems*, vol. 19, no. 6, pp. 1107–1126, 2011.

[207] L. Yang, C. Chen, N. Jin, X. Fu, and Q. Shen, "Closed form fuzzy interpolation with interval type-2 fuzzy sets," in *Proceedings of the IEEE International Conference on Fuzzy Systems*, 2014, pp. 2184–2191.

[208] L. Yang and Q. Shen, "Closed form fuzzy interpolation," *Fuzzy Sets and Systems*, vol. 225, pp. 1–22, 2013.

[209] J. Yen and L. Wang, "Simplifying fuzzy rule-based models using orthogonal transformation methods," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 29, no. 1, pp. 13–24, 1999.

[210] D. S. Yeung and E. C. C. Tsang, "A comparative study on similarity-based fuzzy reasoning methods," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 27, no. 2, pp. 216–227, 1997.

[211] Y. Yuan and H. Zhuang, "A genetic algorithm for generating fuzzy classification rules," *Fuzzy Sets and Systems*, vol. 84, no. 1, pp. 1–19, 1996.

[212] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.

[213] ——, "A fuzzy-set-theoretic interpretation of linguistic hedges," *Cybernetics and Systems*, vol. 2, no. 3, pp. 4–34, 1972.

[214] ——, "Outline of a new approach to the analysis of complex systems and decision processes," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 3, no. 1, pp. 28–44, 1973.

[215] ——, "The concept of a linguistic variable and its application to approximate reasoning–I," *Information Sciences*, vol. 8, no. 3, pp. 199–249, 1975.

[216] ——, "Fuzzy logic and approximate reasoning," *Synthese*, vol. 30, no. 3-4, pp. 407–428, 1975.

[217] ——, "Fuzzy logic," *Computer*, vol. 21, no. 4, pp. 83–93, 1988.

[218] ——, "Knowledge representation in fuzzy logic," *IEEE Transactions on Knowledge and Data Engineering*, vol. 1, no. 1, pp. 89–100, 1989.

[219] ——, "Fuzzy logic, neural networks, and soft computing," *Communications of the ACM*, vol. 37, no. 3, pp. 77–84, 1994.

[220] ——, "Fuzzy logic = computing with words," *IEEE Transactions on Fuzzy Systems*, vol. 4, no. 2, pp. 103–111, 1996.

[221] ——, "Is there a need for fuzzy logic?" *Information Sciences*, vol. 178, no. 13, pp. 2751–2779, 2008.

[222] ——, "Quantitative fuzzy semantics," *Information Sciences*, vol. 3, no. 2, pp. 159–176, 1971.

[223] M. Zeleny and J. L. Cochrane, *Multiple criteria decision making*. McGraw-Hill, 1982.

[224] Q. Zhang, "Fuzziness-vagueness-generality-ambiguity," *Journal of Pragmatics*, vol. 29, no. 1, pp. 13–31, 1998.

[225] H. Zimmerman, *Fuzzy sets, decision making, and expert systems*. Kluwer Academic Publishers, 1987.

[226] ——, *Fuzzy set theory and its applications*. Kluwer Academic Publishers, 1991.